



(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
27.04.2005 Bulletin 2005/17

(51) Int Cl.7: **C12N 15/31, C12N 1/21,
C12P 21/02, C07K 14/33,
A61K 38/16, A61K 39/08**

(21) Application number: **97942094.0**

(86) International application number:
PCT/GB1997/002273

(22) Date of filing: **22.08.1997**

(87) International publication number:
WO 1998/007864 (26.02.1998 Gazette 1998/08)

(54) **RECOMBINANT TOXIN FRAGMENTS**
REKOMBINANTE TOXINFRAGMENTE
FRAGMENTS DE TOXINES RECOMBINES

(84) Designated Contracting States:
AT BE CH DE DK FR GB IT LI NL SE

(56) References cited:
WO-A-96/12802

(30) Priority: **23.08.1996 GB 9617671**
13.12.1996 GB 9625996

(43) Date of publication of application:
08.09.1999 Bulletin 1999/36

(73) Proprietors:
• **Health Protection Agency**
Salisbury, Wiltshire SP4 0JG (GB)
• **Ipsen Limited**
Slough Berkshire SL1 3XE (GB)

(72) Inventors:
• **SHONE, Clifford, Charles, CAMR**
Porton Down, Salisbury, Wiltshire SP4 0JG (GB)
• **QUINN, Conrad, Padraig, CAMR**
Porton Down, Salisbury, Wiltshire SP4 0JG (GB)
• **FOSTER, Keith, Alan, CAMR**
Porton Down, Salisbury, Wiltshire SP4 0JG (GB)

(74) Representative: **MacLean, Martin Robert**
Mathys & Squire
120 Holborn
London EC1N 2SQ (GB)

- **KURAZONO H ET AL: "Minimal essential *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910**
- **LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938**
- **BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348**

Remarks:

The file contains technical information submitted after the application was filed and not included in this specification

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

[0001] This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

[0002] The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

[0003] Inhibition of calcium-independent [^3H] noradrenaline outflow from freeze-thawed synaptosomes has also been reported (Hausinger, A. et al 1995, Toxicon, vol. 33, No. 11, p. 1519 - 1530)

[0004] The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the H_C domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the H_N domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The H_N domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

[0005] For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the H_C , amino acid residues 449-871 for the H_N and residues 1-448 for the LC. Digestion with trypsin effectively degrades the H_C domain of the BoNT/A to generate a non-toxic fragment designated LH_N , which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

[0006] It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

(A) clostridial neurotoxin light chain:

- a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma - membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin. A single mutation (Glu²²⁴ \rightarrow Ala²³⁴) in the light chain of tetanus toxin has been shown to abolish proteolytic activity (Li, Y. et al, Biochemistry 1994, 33, p. 7014 - 7020).

(B) clostridial neurotoxin heavy chain H_N domain:

- a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.

(C) clostridial neurotoxin heavy chain H_C domain.

- a portion of the heavy chain which is responsible for binding of the native holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

[0007] The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

[0008] It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

[0009] For example, WO96/12802 describes vaccine compositions comprising the H_C region of *C. botulinum*.

[0010] However, due to its extreme toxicity, the handling of native toxin is hazardous.

[0011] The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

5 [0012] Accordingly, the invention provides a polypeptide as defined in Claim 1.

[0013] The invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H_C domain.

10 [0014] For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a H_N domain or to the functions of a H_N domain are references to this property or properties. The second domain is not required to exhibit other properties of the H_N domain of a clostridial toxin heavy chain.

15 [0015] A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up

20 of that polypeptide and one or more other components by one or more of said other components.

[0016] The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from *Aplysia* (sea slug), *Drosophila* (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

25 [0017] The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

30 [0018] In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in *J. Biol. Chem.*, Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

40 [0019] In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

45 [0020] The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

50 [0021] In an embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

55 [0022] In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain H_N portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the H_N domain. Teachings of regions within the H_N responsible for its function are provided for example in *Biochemistry* 1995, 34, pages 15175-15181 and *Eur. J. Biochem.* 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment,

though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a H_N domain or fragment, (ii) a C-terminal extension to a H_N domain or fragment, (iii) a modification to a H_N domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

[0023] The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

[0024] These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is significantly reduced in an *in vitro* assay of neurotoxin endopeptidase activity.

[0025] Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

[0026] In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated H_C of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated H_C of a clostridial toxin heavy chain.

[0027] In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C, D, E, F and G.

[0028] The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

[0029] In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

[0030] One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

[0031] As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide.

[0032] A third aspect of the invention provides a composition comprising a polypeptide of the present invention, said composition being non-toxic *in vivo*. The overall endopeptidase activity of the composition will, of course, be determined by the amount of the polypeptide that is present.

[0033] While it is known to treat naturally produced clostridial toxin to remove the H_C domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

[0034] The invention enables production of the polypeptides and fusion proteins of the invention by recombinant

means.

[0035] A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

5 [0036] In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

[0037] A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain.
10 This recombinant product is designated LH₄₂₃/A (SEQ ID NO: 2).

[0038] In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

15 [0039] A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH₇₂₈/B (SEQ ID NO: 20).

[0040] The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

20 [0041] The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

25 [0042] The LH_N/A derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal half of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

30 [0043] In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide, or if expression hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH₄₂₃/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

35 [0044] The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_C domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

45 [0045] One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EPA-0763131.

50 [0046] A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

[0047] In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

55 [0048] The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

[0049] Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

[0050] Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a '/' to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

[0051] Following this nomenclature,

$LH_{423}/A =$ SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;

$2LH_{423}/A =$ a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

$2L_{1/2}H_{423}/A =$ a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain;

$2L_{FXa/2}H_{423}/A =$ a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

$2L_{FXa/2}H_{423}/A\text{-IGF-1} =$ a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

[0052] There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH_{423}/A ;

Fig.3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designated $2LH_{423}/A(Q_2E, N_{26}K, A_{27}Y)$ in an *in vitro* peptide cleavage assay;

Fig.4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;

Fig.5 shows the transition region of an embodiment of the invention designated L_4H_{423}/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;

Fig.6 shows the transition region of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;

Fig.7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A\text{-IGF-1}$, a fusion protein; the IGF-1 sequence begins at position G₈₈₂;

Fig.8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated

EP 0 939 818 B1

L_{FXa/3}H₄₂₃/A-CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Q₈₈₂;

Fig.9 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated L_{FXa/3}H₄₂₃/A-ZZ, a fusion protein; the C-terminal ZZ sequence begins at position A₈₉₀ immediately after a genesase recognition site (underlined);

Figs.10 & 11 show schematic representations of manipulations of polypeptides of the invention; Fig.10 shows LH₄₂₃/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig.12 shows the trypsin sensitive activation region of a polypeptide of the invention;

Fig.13 shows Western blot analysis of recombinant LH₁₀₇/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH₁₀₇/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH₁₀₇/B.

[0053] The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:	SEQUENCE
1	DNA coding for LH ₄₂₃ /A
2	LH ₄₂₃ /A
3	DNA coding for ₂₃ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an N-terminal portion is shown in Fig. 4.
4	₂₃ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
5	DNA coding for ₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an N-terminal portion is shown in Fig.4
6	₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
7	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L _{1/4} H ₄₂₃ /A
10	L _{1/4} H ₄₂₃ /A
11	DNA coding for L _{FXa/3} H ₄₂₃ /A
12	L _{FXa/3} H ₄₂₃ /A
13	DNA coding for L _{FXa/3} H ₄₂₃ /A-IGF-1
14	L _{FXa/3} H ₄₂₃ /A-IGF-1
15	DNA coding for L _{FXa/3} H ₄₂₃ /A-CtxA14
16	L _{FXa/3} H ₄₂₃ /A-CtxA 14
17	DNA coding for L _{FXa/3} H ₄₂₃ /A-ZZ
18	L _{FXa/3} H ₄₂₃ /A-ZZ
19	DNA coding for LH ₇₂₈ /B
20	LH ₇₂₈ /B
21	DNA coding for LH ₄₁₇ /B
22	LH ₄₁₇ /B
23	DNA coding for LH ₁₀₇ /B
24	LH ₁₀₇ /B
25	DNA coding for LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
26	LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
27	DNA coding for LH ₄₁₇ /B wherein the first 274 bases are modified to have an <i>E.coli</i> codon bias
28	DNA coding for LH ₄₁₇ /B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

[0054] A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_C) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

[0055] The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an *E. coli* codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique *KpnI* restriction site. The remainder of the LH₄₂₃/A coding sequence was PCR amplified from total chromosomal DNA from *Clostridium botulinum* and annealed to the synthetic portion of the gene.

[0056] The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The assembled gene was then fully sequenced and cloned into a variety of *E. coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

[0057] The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

[0058] Currently, *E. coli* harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced by addition of isopropylthio- β -D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH₄₂₃/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

[0059] The recombinant GST- LH₄₂₃/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH₄₂₃/A is recovered in the non-adsorbed material.

Construct variants

[0060] A variant of the molecule, LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH₄₂₃/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

[0061] Two further variants of the gene sequence that have been expressed and the corresponding products purified are ₂₃LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

[0062] In yet another variant a gene has been produced which contains a *Eco* 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the

capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the *Eco* 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, $L_{1/4}H_{423}/A$ (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of LH_{423}/A at a position equivalent to the amino terminus of the heavy chain of native BoNT/A.

[0063] A variant of the gene has been expressed, $L_{FXa/3}H_{423}/A$ (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of $L_{1/4}H_{423}/A$. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

[0064] Variants of $L_{FXa/3}H_{423}/A$ have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

[0065] Specific examples described are:

(1) $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;

(2) $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and

(3) $L_{FXa/3}H_{423}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

[0066] It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

[0067] Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

[0068] The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

[0069] Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H_N domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

[0070] The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The $_2\text{LH}_{423}/\text{A}$ ($\text{Q}_2\text{E}, \text{N}_{26}\text{K}, \text{A}_{27}\text{Y}$) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of $_2\text{LH}_{423}/\text{A}$ ($\text{Q}_2\text{E}, \text{N}_{26}\text{K}, \text{A}_{27}\text{Y}$) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

[0071] This activity is dependent on proteolytic modification of the recombinant GST- $_2\text{LH}_{423}/\text{A}$ ($\text{Q}_2\text{E}, \text{N}_{26}\text{K}, \text{A}_{27}\text{Y}$) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 $\mu\text{g}/\text{ml}$) is incubated at 37°C for 10-50 minutes with trypsin (10 $\mu\text{g}/\text{ml}$) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , 1.8 mM KH_2PO_4 , pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

[0072] $_2\text{LH}_{423}/\text{A}$ is more stable in the presence of trypsin and more active in the *in vitro* peptide cleavage assay than is $_{23}\text{LH}_{423}/\text{A}$. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

[0073] As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

[0074] The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

[0075] A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH_{728}/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

[0076] A gene has also been assembled coding for a variant polypeptide, LH_{417}/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native LH_N/A .

[0077] A gene has also been assembled coding for a variant polypeptide, LH_{107}/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

[0078] A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

[0079] Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E. coli* codon usage bias.

[0080] For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27).

Similarly the sequence could be inserted into other genes of the examples.

EP 0 939 818 B1

[0081] Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

[0082] An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-₂LH₄₂₃/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA 11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

[0083] For immunisation, 20 µg of GST-₂LH₄₂₃/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH_N/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na₂HPO₄ anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 µg (5X10⁻⁶ g) to 5 picograms (5X10⁻¹² g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- GST-₂LH₄₂₃/A antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1.

Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST- ₂ LH ₄₂₃ /A antiserum.							
Botulinum Toxin/mouse							
Survivors On Day	0.5µg	0.005µg	0.0005pg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	4	4	4	4	4	4
2	-	4	4	4	4	4	4
3	-	4	4	4	4	4	4
4	-	4	4	4	4	4	4

TABLE 2.

Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.							
Botulinum Toxin/mouse							
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	0	0	0	0	2	4
2	-	-	-	-	-	0	4
3	-	-	-	-	-	-	4
4	-	-	-	-	-	-	4

Example 4

Expression of recombinant LH₁₀₇/B in *E. coli*.

[0084] As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide

EP 0 939 818 B1

LH₁₀₇/B (SEQ ID NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in *E. coli* BL21 (DE3) (New England BioLabs, Beverly, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for LH₄₂₃/A.

[0085] Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverly, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

[0086] The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

[0087] The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

SEQUENCE LISTING

[0088]

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
(B) STREET: Centre For Applied Microbiology And Research, Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(A) NAME: THE SPEYWOOD LABORATORY LIMITED
(B) STREET: 14 Kensington Square
(C) CITY: London
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): W8 5HH

(A) NAME: FOSTER; Keith Alan
(B) STREET: Centre For Applied Microbiology And Research, Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(A) NAME: QUINN; Conrad Padraig
(B) STREET: Centre For Applied Microbiology And Research, Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(A) NAME: SHONE; Clifford Charles
(B) STREET: Centre For Applied Microbiology And Research, Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

EP 0 939 818 B1

(ii) TITLE OF INVENTION: Recombinant Toxin Fragments

(iii) NUMBER OF SEQUENCES: 28

5

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

10

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2616 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

25

(A) NAME/KEY: CDS

(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

30

35

40

45

50

55

EP 0 939 818 B1

	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15	48
5	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30	96
10	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60	192
15	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
20	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
25	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
30	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
35	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
40	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
45	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
50	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720

EP 0 939 818 B1

	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA	768
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu	
	245 250 255	
5	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG	816
	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys	
	260 265 270	
10	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC	864
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn	
	275 280 285	
	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG	912
	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val	
	290 295 300	
15	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA	960
	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys	
	305 310 315 320	
20	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA	1008
	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu	
	325 330 335	
	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT	1056
	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp	
	340 345 350	
25	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT	1104
	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn	
	355 360 365	
	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC	1152
	Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr	
	370 375 380	
30	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC	1200
	Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn	
	385 390 395 400	
35	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA	1248
	Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu	
	405 410 415	
	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA	1296
	Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg	
	420 425 430	
40	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG	1344
	Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys	
	435 440 445	
	GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT	1392
	Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe	
	450 455 460	
45	AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA	1440
	Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu	
	465 470 475 480	
50	ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT AGT TTA	1488
	Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu	
	485 490 495	
	GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT	1536
	Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro	
	500 505 510	

EP 0 939 818 B1

	GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
5	GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
10	TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560	1680
	CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
15	TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
20	AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
	CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
25	GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640	1920
	TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT GTA GGT GCT TTA Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655	1968
30	ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
35	ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
	GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
40	AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720	2160
	GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735	2208
45	GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750	2256
50	CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT GAT GAT Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765	2304
55	TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780	2352

EP 0 939 818 B1

5	AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT TCT ATG Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800	2400
	ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT CTT AAA Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815	2448
10	GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA ATT GGT Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830	2496
15	CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT ACA GAT Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845	2544
	ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA TTA TCT Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860	2592
20	ACA TTT ACT GAA TAT ATT AAG TAA Thr Phe Thr Glu Tyr Ile Lys * 865 870	2616

(2) INFORMATION FOR SEQ ID NO: 2:

- 25 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- 30 (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
- 35
- 40
- 45
- 50
- 55

EP 0 939 818 B1

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15
5 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45
10 Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50 55 60
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80
15 Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
85 90 95
Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
100 105 110
20 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125
Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140
25 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145 150 155 160
Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
165 170 175

30

35

40

45

50

55

EP 0 939 818 B1

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190
 5 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205
 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220
 10 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240
 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255
 15 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 20 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 25 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 30 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 35 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 40 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 45 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
 450 455 460
 Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480
 50 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 55 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525

EP 0 939 818 B1

	Glu	Leu	Met	Pro	Asn	Ile	Glu	Arg	Phe	Pro	Asn	Gly	Lys	Lys	Tyr	Glu	
	530						535					540					
5	Leu	Asp	Lys	Tyr	Thr	Met	Phe	His	Tyr	Leu	Arg	Ala	Gln	Glu	Phe	Glu	
	545					550					555					560	
	His	Gly	Lys	Ser	Arg	Ile	Ala	Leu	Thr	Asn	Ser	Val	Asn	Glu	Ala	Leu	
					565					570					575		
10	Leu	Asn	Pro	Ser	Arg	Val	Tyr	Thr	Phe	Phe	Ser	Ser	Asp	Tyr	Val	Lys	
			580						585					590			
	Lys	Val	Asn	Lys	Ala	Thr	Glu	Ala	Ala	Met	Phe	Leu	Gly	Trp	Val	Glu	
			595					600					605				
15	Gln	Leu	Val	Tyr	Asp	Phe	Thr	Asp	Glu	Thr	Ser	Glu	Val	Ser	Thr	Thr	
	610						615					620					
	Asp	Lys	Ile	Ala	Asp	Ile	Thr	Ile	Ile	Ile	Pro	Tyr	Ile	Gly	Pro	Ala	
	625					630					635					640	
20	Leu	Asn	Ile	Gly	Asn	Met	Leu	Tyr	Lys	Asp	Asp	Phe	Val	Gly	Ala	Leu	
					645					650					655		
	Ile	Phe	Ser	Gly	Ala	Val	Ile	Leu	Leu	Glu	Phe	Ile	Pro	Glu	Ile	Ala	
				660					665					670			
25	Ile	Pro	Val	Leu	Gly	Thr	Phe	Ala	Leu	Val	Ser	Tyr	Ile	Ala	Asn	Lys	
			675					680					685				
	Val	Leu	Thr	Val	Gln	Thr	Ile	Asp	Asn	Ala	Leu	Ser	Lys	Arg	Asn	Glu	
		690					695					700					
30	Lys	Trp	Asp	Glu	Val	Tyr	Lys	Tyr	Ile	Val	Thr	Asn	Trp	Leu	Ala	Lys	
	705					710					715					720	
	Val	Asn	Thr	Gln	Ile	Asp	Leu	Ile	Arg	Lys	Lys	Met	Lys	Glu	Ala	Leu	
				725						730					735		
35	Glu	Asn	Gln	Ala	Glu	Ala	Thr	Lys	Ala	Ile	Ile	Asn	Tyr	Gln	Tyr	Asn	
				740					745					750			
	Gln	Tyr	Thr	Glu	Glu	Glu	Lys	Asn	Asn	Ile	Asn	Phe	Asn	Ile	Asp	Asp	
			755					760					765				
40	Leu	Ser	Ser	Lys	Leu	Asn	Glu	Ser	Ile	Asn	Lys	Ala	Met	Ile	Asn	Ile	
		770				775						780					
	Asn	Lys	Phe	Leu	Asn	Gln	Cys	Ser	Val	Ser	Tyr	Leu	Met	Asn	Ser	Met	
	785					790					795					800	
45	Ile	Pro	Tyr	Gly	Val	Lys	Arg	Leu	Glu	Asp	Phe	Asp	Ala	Ser	Leu	Lys	
				805						810					815		
	Asp	Ala	Leu	Leu	Lys	Tyr	Ile	Tyr	Asp	Asn	Arg	Gly	Thr	Leu	Ile	Gly	
				820					825					830			
50	Gln	Val	Asp	Arg	Leu	Lys	Asp	Lys	Val	Asn	Asn	Thr	Leu	Ser	Thr	Asp	
			835					840					845				
	Ile	Pro	Phe	Gln	Leu	Ser	Lys	Tyr	Val	Asp	Asn	Gln	Arg	Leu	Leu	Ser	
				850			855					860					
55	Thr	Phe	Thr	Glu	Tyr	Ile	Lys	*									
	865					870											

(2) INFORMATION FOR SEQ ID NO: 3:

EP 0 939 818 B1

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 2685 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION:1..2685

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

20

25

30

35

40

45

50

55

EP 0 939 818 B1

	GGA TCC CCA GGA ATT CAT ATG ACG TCG ACG CGT CTG CAG AAG CTT CTA Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu 1 5 10 15	48
5	GAA TTC GAG CTC CCG GGT ACC ATG GAG TTC GTG AAC AAG CAG TTC AAC Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30	96
10	TAT AAG GAC CCT GTA AAC GGT GTT GAC ATT GCC TAC ATC AAA ATT CCA Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro 35 40 45	144
	AAG TAC GGC CAG ATG CAG CCG GTG AAG GCT TTC AAG ATT CAT AAC AAA Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60	192
15	ATC TGG GTT ATT CCG GAA CGC GAT ACA TTT ACG AAC CCG GAA GAA GGA Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80	240
20	GAC TTG AAC CCG CCG CCG GAA GCA AAG CAG GTG CCA GTT TCA TAC TAC Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95	288
	GAT TCA ACC TAT CTG AGC ACA GAC AAC GAG AAG GAT AAC TAC CTG AAG Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys 100 105 110	336
25	GGA GTG ACC AAA TTA TTC GAG CGT ATT TAT TCC ACT GAC CTG GGC CGT Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg 115 120 125	384
30	ATG CTG CTG ACC TCA ATC GTC CGC GGA ATC CCA TTT TGG GGT GGC AGT Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser 130 135 140	432
	ACC ATT GAC ACG GAG TTG AAG GTT ATT GAC ACT AAC TGC ATT AAC GTG Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val 145 150 155 160	480
35	ATC CAA CCA GAC GGT AGC TAC AGA TCT GAA GAA CTT AAC CTC GTA ATC Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile 165 170 175	528
40	ATC GGG CCC TCC GCG GAC ATT ATC CAG TTT GAG TGC AAG AGC TTT GGC Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly 180 185 190	576
45	CAC GAA GTG TTG AAC CTG ACG CGT AAC GGT TAC GGC TCT ACT CAG TAC His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr 195 200 205	624

EP 0 939 818 B1

	ATT CGT TTC AGC CCA GAC TTC ACG TTC GGT TTC GAG GAG AGC CTG GAG Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu 210 215 220	672
5	GTT GAT ACC AAC CCG CTG TTG GGT GCA GGC AAG TTC GCA ACT GAT CCA Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro 225 230 235 240	720
10	GCG GTG ACC CTG GCA CAC GAG CTG ATC CAC GCC GGT CAT CGT CTG TAT Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr 245 250 255	768
	GGC ATT GCG ATT AAC CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 260 265 270	816
15	TAC TAC GAG ATG AGT GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACG Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr 275 280 285	864
20	TTC GGT GGC CAT GAT GCG AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu 290 295 300	912
	TTC CGT CTG TAC TAC TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu 305 310 315 320	960
25	AAC AAG GCT AAG TCC ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met 325 330 335	1008
30	AAA AAT GTT TTT AAA GAG AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly 340 345 350	1056
	AAA TTT TCG GTA GAT AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu 355 360 365	1104
35	ACA GAG ATT TAC ACA GAG GAT AAT TTT GTT AAG TTT TTT AAA GTA CTT Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu 370 375 380	1152
	AAC AGA AAA ACA TAT TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn 385 390 395 400	1200
40	ATA GTA CCT AAG GTA AAT TAC ACA ATA TAT GAT GGA TTT AAT TTA AGA Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg 405 410 415	1248
45	AAT ACA AAT TTA GCA GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn 420 425 430	1296
	AAT ATG AAT TTT ACT AAA CTA AAA AAT TTT ACT GGA TTG TTT GAA TTT Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe 435 440 445	1344
50	TAT AAG TTG CTA TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser 450 455 460	1392
55	TTA GAT AAA GGA TAC AAT AAG GCA TTA AAT GAT TTA TGT ATC AAA GTT Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 465 470 475 480	1440

EP 0 939 818 B1

	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 485 490 495	1488
5	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 500 505 510	1536
10	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 515 520 525	1584
	TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 530 535 540	1632
15	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 545 550 555 560	1680
20	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 565 570 575	1728
	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 580 585 590	1776
25	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 595 600 605	1824
30	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 610 615 620	1872
	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 625 630 635 640	1920
35	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 645 650 655	1968
	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 660 665 670	2016
40	GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 675 680 685	2064
45	GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 690 695 700	2112
	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 705 710 715 720	2160
50	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 725 730 735	2208
55	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 740 745 750	2256

EP 0 939 818 B1

5	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 755 760 765	2304
10	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 770 775 780	2352
15	ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 785 790 795 800	2400
20	AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 805 810 815	2448
25	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 820 825 830	2496
30	GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 835 840 845	2544
35	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 850 855 860	2592
40	AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 865 870 875 880	2640
45	GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 885 890 895	2685

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 895 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

EP 0 939 818 B1

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu
 1 5 10 15
 Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn
 5 20 25 30 —
 Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro
 35 40 45
 Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys
 10 50 55 60
 Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly
 65 70 75 80
 Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr
 15 85 90 95
 Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys
 100 105 110

20

25

30

35

40

45

50

55

EP 0 939 818 B1

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg
 115 120 125
 5 Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser
 130 135 140
 Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val
 145 150 155 160
 10 Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile
 165 170 175
 Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly
 180 185 190
 15 His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr
 195 200 205
 Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu
 210 215 220
 20 Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro
 225 230 235 240
 Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr
 245 250 255
 25 Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala
 260 265 270
 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr
 275 280 285
 30 Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu
 290 295 300
 Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu
 305 310 315 320
 35 Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met
 325 330 335
 Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly
 340 345 350
 40 Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu
 355 360 365
 Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu
 370 375 380
 45 Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn
 385 390 395 400
 Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg
 405 410 415
 50 Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn
 420 425 430
 Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe
 435 440 445
 55 Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser
 450 455 460

EP 0 939 818 B1

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val
 465 470 475 480
 5 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 485 490 495
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 500 505 510
 10 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 515 520 525
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 530 535 540
 15 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 545 550 555 560
 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 565 570 575
 20 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 580 585 590
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 595 600 605
 25 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 610 615 620
 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 625 630 635 640
 30 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 645 650 655
 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 660 665 670
 35 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 675 680 685
 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 690 695 700
 40 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 705 710 715 720
 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 725 730 735
 45 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 740 745 750
 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 755 760 765
 50 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 770 775 780
 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 785 790 795 800
 55 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 805 810 815

EP 0 939 818 B1

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
820 825 830
5 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
835 840 845
Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
850 855 860
10 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
865 870 875 880
Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys +
885 890 895

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2622

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

EP 0 939 818 B1

	GGA TCC ATG GAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val 1 5 10 15	48
5	AAC GGT GTT GAC ATT GCC TAC ATC AAA ATT CCA AAG TAC GGC CAG ATG Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30	96
10	CAG CCG GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro 35 40 45	144
	GAA CGC GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro 50 55 60	192
15	CCG GAA GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu 65 70 75 80	240
20	AGC ACA GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu 85 90 95	288
	TTC GAG CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser 100 105 110	336
25	ATC GTC CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu 115 120 125	384
30	TTG AAG GTT ATT GAC ACT AAC TGC ATT AAC CTG ATC CAA CCA GAC GGT Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly 130 135 140	432
35		
40		
45		
50		
55		

EP 0 939 818 B1

	AGC TAC AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala 145 150 155 160	480
5	GAC ATT ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn 165 170 175	528
10	CTG ACG CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro 180 185 190	576
	GAC TTC ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 195 200 205	624
15	CTG TTG GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala 210 215 220	672
20	CAC GAG CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn 225 230 235 240	720
	CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 245 250 255	768
25	GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 260 265 270	816
30	GCG AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 275 280 285	864
	TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 290 295 300	912
35	ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 305 310 315 320	960
	GAG AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 325 330 335	1008
40	AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 345 350	1056
45	GAG GAT AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr 355 360 365	1104
	TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 370 375 380	1152
50	AAT TAC ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 385 390 395 400	1200
55	GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 405 410 415	1248

EP 0 939 818 B1

	AAA CTA AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT	1296
	Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys	
	420 425 430	
5	GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC	1344
	Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr	
	435 440 445	
10	AAT AAG GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG	1392
	Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu	
	450 455 460	
	TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA	1440
	Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly	
	465 470 475 480	
15	GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT	1488
	Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile	
	485 490 495	
20	AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT	1536
	Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn	
	500 505 510	
	GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC	1584
	Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly	
	515 520 525	
25	CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG	1632
	Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys	
	530 535 540	
30	TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA	1680
	Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu	
	545 550 555 560	
	TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA	1728
	Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu	
	565 570 575	
35	GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT	1776
	Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr	
	580 585 590	
	GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG	1824
	Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp	
	595 600 605	
40	GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT	1872
	Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser	
	610 615 620	
45	ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA	1920
	Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly	
	625 630 635 640	
	CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT GTA GGT	1968
	Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly	
	645 650 655	
50	GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG	2016
	Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu	
	660 665 670	
55	ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG	2064
	Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala	
	675 680 685	

EP 0 939 818 B1

	AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg 690 695 700	2112
5	AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu 705 710 715 720	2160
10	GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu 725 730 735	2208
	GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln 740 745 750	2256
15	TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile 755 760 765	2304
20	GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile 770 775 780	2352
	AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn 785 790 795 800	2400
25	TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser 805 810 815	2448
30	CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu 820 825 830	2496
	ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser 835 840 845	2544
35	ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu 850 855 860	2592
40	TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 865 870	2622

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 874 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

EP 0 939 818 B1

Gly	Ser	Met	Glu	Phe	Val	Asn	Lys	Gln	Phe	Asn	Tyr	Lys	Asp	Pro	Val
1				5					10					15	
Asn	Gly	Val	Asp	Ile	Ala	Tyr	Ile	Lys	Ile	Pro	Lys	Tyr	Gly	Gln	Met
			20					25					30		
Gln	Pro	Val	Lys	Ala	Phe	Lys	Ile	His	Asn	Lys	Ile	Trp	Val	Ile	Pro
		35					40					45			

10

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro
 50 55 60
 5 Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu
 65 70 75 80
 Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu
 85 90 95
 10 Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser
 100 105 110
 Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu
 115 120 125
 15 Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly
 130 135 140
 Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala
 145 150 155 160
 20 Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn
 165 170 175
 Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro
 180 185 190
 25 Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro
 195 200 205
 Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala
 210 215 220
 30 His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn
 225 230 235 240
 Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser
 245 250 255
 35 Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp
 260 265 270
 Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr
 275 280 285
 40 Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser
 290 295 300
 Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys
 305 310 315 320
 45 Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp
 325 330 335
 Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr
 340 345 350
 50 Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr
 355 360 365
 Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val
 370 375 380
 55 Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala
 385 390 395 400

EP 0 939 818 B1

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr
405 410 415

5 Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys
420 425 430

Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr
435 440 445

10 Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu
450 455 460

Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly
465 470 475 480

15 Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile
485 490 495

Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn
500 505 510

20 Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly
515 520 525

Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys
530 535 540

25 Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu
545 550 555 560

Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu
565 570 575

30 Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr
580 585 590

Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp
595 600 605

35 Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser
610 615 620

Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly
625 630 635 640

40 Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly
645 650 655

Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu
660 665 670

45 Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala
675 680 685

Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg
690 695 700

50 Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu
705 710 715 720

Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu
725 730 735

55 Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln
740 745 750

EP 0 939 818 B1

Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile
 755 760 765
 5 Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile
 770 775 780
 Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn
 785 790 795 800
 10 Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser
 805 810 815
 Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu
 820 825 830
 15 Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser
 835 840 845
 Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu
 850 855 860
 20 Leu Ser Thr Phe Thr Glu Tyr Ile Lys *
 865 870

(2) INFORMATION FOR SEQ ID NO: 7:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2613 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2613

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

EP 0 939 818 B1

	ATG CCA TTT GTT AAT AAA CAA TTT AAT TAT AAA GAT CCT GTA AAT GGT	48
	Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
	1 5 10 15	
5	GTT GAT ATT GCT TAT ATA AAA ATT CCA AAT GCA GGA CAA ATG CAA CCA	96
	Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
	20 25 30	
10	GTA AAA GCT TTT AAA ATT CAT AAT AAA ATA TGG GTT ATT CCA GAA AGA	144
	Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
	35 40 45	
15	GAT ACA TTT ACA AAT CCT GAA GAA GGA GAT TTA AAT CCA CCA CCA GAA	192
	Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	
	50 55 60	
20	GCA AAA CAA GTT CCA GTT TCA TAT TAT GAT TCA ACA TAT TTA AGT ACA	240
	Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	
	65 70 75 80	
25	GAT AAT GAA AAA GAT AAT TAT TTA AAG GGA GTT ACA AAA TTA TTT GAG	288
	Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	
	85 90 95	

EP 0 939 818 B1

	AGA ATT TAT TCA ACT GAT CTT GGA AGA ATG TTG TTA ACA TCA ATA GTA	336
	Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	
	100 105 110	
5	AGG GGA ATA CCA TTT TGG GGT GGA AGT ACA ATA GAT ACA GAA TTA AAA	384
	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	
	115 120 125	
10	GTT ATT GAT ACT AAT TGT ATT AAT GTG ATA CAA CCA GAT GGT AGT TAT	432
	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr	
	130 135 140	
	AGA TCA GAA GAA CTT AAT CTA GTA ATA ATA GGA CCC TCA GCT GAT ATT	480
	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile	
	145 150 155 160	
15	ATA CAG TTT GAA TGT AAA AGC TTT GGA CAT GAA GTT TTG AAT CTT ACG	528
	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr	
	165 170 175	
20	CGA AAT GGT TAT GGC TCT ACT CAA TAC ATT AGA TTT AGC CCA GAT TTT	576
	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	
	180 185 190	
	ACA TTT GGT TTT GAG GAG TCA CTT GAA GTT GAT ACA AAT CCT CTT TTA	624
	Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu	
	195 200 205	
25	GGT GCA GGC AAA TTT GCT ACA GAT CCA GCA GTA ACA TTA GCA CAT GAA	672
	Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu	
	210 215 220	
30	CTT ATA CAT GCT GGA CAT AGA TTA TAT GGA ATA GCA ATT AAT CCA AAT	720
	Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn	
	225 230 235 240	
	AGG GTT TTT AAA GTA AAT ACT AAT GCC TAT TAT GAA ATG AGT GGG TTA	768
	Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu	
	245 250 255	
35	GAA GTA AGC TTT GAG GAA CTT AGA ACA TTT GGG GGA CAT GAT GCA AAG	816
	Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys	
	260 265 270	
	TTT ATA GAT AGT TTA CAG GAA AAC GAA TTT CGT CTA TAT TAT TAT AAT	864
	Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn	
	275 280 285	
40	AAG TTT AAA GAT ATA GCA AGT ACA CTT AAT AAA GCT AAA TCA ATA GTA	912
	Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val	
	290 295 300	
45	GGT ACT ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA	960
	Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys	
	305 310 315 320	
	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA	1008
	Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu	
	325 330 335	
50	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT	1056
	Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp	
	340 345 350	
55	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT	1104
	Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn	
	355 360 365	

EP 0 939 818 B1

	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
5	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
10	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
15	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
20	GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 455 460	1392
	AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480	1440
25	ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT AGT TTA Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	1488
30	GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510	1536
	GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
35	GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
40	TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560	1680
	CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
45	TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
	AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
50	CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
55	GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640	1920

EP 0 939 818 B1

	TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT GTA GGT GCT TTA Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655	1968
5	ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
10	ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
	GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
15	AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720	2160
20	GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735	2208
	GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750	2256
25	CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT GAT GAT Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765	2304
30	TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780	2352
	AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT TCT ATG Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800	2400
35	ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT CTT AAA Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815	2448
	GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA ATT GGT Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830	2496
40	CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT ACA GAT Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845	2544
45	ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA TTA TCT Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860	2592
	ACA TTT ACT GAA TAT ATT AAG Thr Phe Thr Glu Tyr Ile Lys 865 870	2613

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 871 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

EP 0 939 818 B1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

5	Met	Pro	Phe	Val	Asn	Lys	Gln	Phe	Asn	Tyr	Lys	Asp	Pro	Val	Asn	Gly	1	5	10	15
	Val	Asp	Ile	Ala	Tyr	Ile	Lys	Ile	Pro	Asn	Ala	Gly	Gln	Met	Gln	Pro	20	25	30	
10	Val	Lys	Ala	Phe	Lys	Ile	His	Asn	Lys	Ile	Trp	Val	Ile	Pro	Glu	Arg	35	40	45	
	Asp	Thr	Phe	Thr	Asn	Pro	Glu	Glu	Gly	Asp	Leu	Asn	Pro	Pro	Pro	Glu	50	55	60	
15	Ala	Lys	Gln	Val	Pro	Val	Ser	Tyr	Tyr	Asp	Ser	Thr	Tyr	Leu	Ser	Thr	65	70	75	80
	Asp	Asn	Glu	Lys	Asp	Asn	Tyr	Leu	Lys	Gly	Val	Thr	Lys	Leu	Phe	Glu	85	90	95	
20	Arg	Ile	Tyr	Ser	Thr	Asp	Leu	Gly	Arg	Met	Leu	Leu	Thr	Ser	Ile	Val	100	105	110	
	Arg	Gly	Ile	Pro	Phe	Trp	Gly	Gly	Ser	Thr	Ile	Asp	Thr	Glu	Leu	Lys	115	120	125	
25	Val	Ile	Asp	Thr	Asn	Cys	Ile	Asn	Val	Ile	Gln	Pro	Asp	Gly	Ser	Tyr	130	135	140	
	Arg	Ser	Glu	Glu	Leu	Asn	Leu	Val	Ile	Ile	Gly	Pro	Ser	Ala	Asp	Ile	145	150	155	160
30	Ile	Gln	Phe	Glu	Cys	Lys	Ser	Phe	Gly	His	Glu	Val	Leu	Asn	Leu	Thr	165	170	175	
	Arg	Asn	Gly	Tyr	Gly	Ser	Thr	Gln	Tyr	Ile	Arg	Phe	Ser	Pro	Asp	Phe	180	185	190	
35	Thr	Phe	Gly	Phe	Glu	Glu	Ser	Leu	Glu	Val	Asp	Thr	Asn	Pro	Leu	Leu	195	200	205	
	Gly	Ala	Gly	Lys	Phe	Ala	Thr	Asp	Pro	Ala	Val	Thr	Leu	Ala	His	Glu	210	215	220	
40	Leu	Ile	His	Ala	Gly	His	Arg	Leu	Tyr	Gly	Ile	Ala	Ile	Asn	Pro	Asn	225	230	235	240
	Arg	Val	Phe	Lys	Val	Asn	Thr	Asn	Ala	Tyr	Tyr	Glu	Met	Ser	Gly	Leu	245	250	255	
45	Glu	Val	Ser	Phe	Glu	Glu	Leu	Arg	Thr	Phe	Gly	Gly	His	Asp	Ala	Lys	260	265	270	
	Phe	Ile	Asp	Ser	Leu	Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	Tyr	Tyr	Asn	275	280	285	
50	Lys	Phe	Lys	Asp	Ile	Ala	Ser	Thr	Leu	Asn	Lys	Ala	Lys	Ser	Ile	Val	290	295	300	
	Gly	Thr	Thr	Ala	Ser	Leu	Gln	Tyr	Met	Lys	Asn	Val	Phe	Lys	Glu	Lys	305	310	315	320
55	Tyr	Leu	Leu	Ser	Glu	Asp	Thr	Ser	Gly	Lys	Phe	Ser	Val	Asp	Lys	Leu	325	330	335	

EP 0 939 818 B1

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 5 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 10 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 15 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 20 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
 450 455 460
 Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480
 25 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 30 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525
 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540
 35 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560
 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575
 40 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590
 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605
 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620
 45 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640
 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655
 50 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670
 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685

EP 0 939 818 B1

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
690 695 700

5 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
725 730 735

10 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
755 760 765

15 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
785 790 795 800

20 Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
820 825 830

25 Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
850 855 860

30 Thr Phe Thr Glu Tyr Ile Lys
865 870

(2) INFORMATION FOR SEQ ID NO: 9:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2628 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2628

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

EP 0 939 818 B1

	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT	48
	Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
	1 5 10 15	
5	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG	96
	Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
	20 25 30	
	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC	144
	Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
	35 40 45	
10	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA	192
	Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	
	50 55 60	
15		
20		
25		
30		
35		
40		
45		
50		
55		

EP 0 939 818 B1

	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
5	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
10	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
15	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
20	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
25	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
30	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
35	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
40	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
45	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
50	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
55	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008

EP 0 939 818 B1

	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
5	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
10	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
15	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
20	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
25	AGC GCT GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp 450 455 460	1392
30	GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480	1440
	AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 495	1488
35	AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 505 510	1536
	GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile 515 520 525	1584
40	ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly 530 535 540	1632
45	AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 550 555 560	1680
	CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 570 575	1728
50	AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser 580 585 590	1776
55	GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 600 605	1824

EP 0 939 818 B1

	GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 615 620	1872
5	GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640	1920
10	ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650 655	1968
	GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile 660 665 670	2016
15	CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr 675 680 685	2064
20	ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser 690 695 700	2112
	AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn 705 710 715 720	2160
25	TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met 725 730 735	2208
	AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn 740 745 750	2256
30	TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe 755 760 765	2304
35	AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala 770 775 780	2352
	ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu 785 790 795 800	2400
40	ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp 805 810 815	2448
45	GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820 825 830	2496
	ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr 835 840 845	2544
50	CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln 850 855 860	2592
55	AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 865 870 875	2628

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 876 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

5 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

10 Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80

15 Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
100 105 110

20 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140

25 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
165 170 175

30 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
195 200 205

35 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
225 230 235 240

40 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
260 265 270

45 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
290 295 300

50

55

EP 0 939 818 B1

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 5 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 10 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 15 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 20 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp
 25 450 455 460
 Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn
 465 470 475 480
 Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu
 30 485 490 495
 Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe
 500 505 510
 Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile
 35 515 520 525
 Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly
 530 535 540
 Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala
 545 550 555 560
 Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val
 40 565 570 575
 Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser
 580 585 590
 Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu
 45 595 600 605
 Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu
 610 615 620
 Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr
 50 625 630 635 640
 Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe
 645 650 655

EP 0 939 818 B1

Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile
660 665 670

5 Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr
675 680 685

Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser
690 695 700

10 Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn
705 710 715 720

Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met
725 730 735

15 Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn
740 745 750

Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe
755 760 765

20 Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala
770 775 780

Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu
785 790 795 800

25 Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp
805 810 815

Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly
820 825 830

30 Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr
835 840 845

Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln
850 855 860

35 Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys *
865 870 875

(2) INFORMATION FOR SEQ ID NO: 11:

40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2637 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
45 (D) TOPOLOGY: linear

(ii) MOLECULE_TYPE: DNA (genomic)

(ix) FEATURE:

50 (A) NAME/KEY: CDS
(B) LOCATION:1..2637

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

55

EP 0 939 818 B1

	ATG	CAG	TTC	GTG	AAC	AAG	CAG	TTC	AAC	TAT	AAG	GAC	CCT	GTA	AAC	GGT	48
	Met	Gln	Phe	Val	Asn	Lys	Gln	Phe	Asn	Tyr	Lys	Asp	Pro	Val	Asn	Gly	
	1				5					10					15		

5	GTT	GAC	ATT	GCC	TAC	ATC	AAA	ATT	CCA	AAC	GCC	GGC	CAG	ATG	CAG	CCG	96
	Val	Asp	Ile	Ala	Tyr	Ile	Lys	Ile	Pro	Asn	Ala	Gly	Gln	Met	Gln	Pro	
				20					25					30			

10

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
5	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60	192
10	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
15	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
20	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
25	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
30	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
35	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
40	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
45	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
50	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
55	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912

EP 0 939 818 B1

	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
5	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
10	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
15	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
20	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
25	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
30	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
35	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
40	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
45	TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632
50	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680
55	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728

EP 0 939 818 B1

	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
5	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
10	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920
15	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
20	GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
	GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
25	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
30	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
35	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
40	ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
45	AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
50	GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
55	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544

EP 0 939 818 B1

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA 2592
 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860

5

GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA 2637
 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys *
 865 870 875

10

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 879 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25

30

35

40

45

50

55

EP 0 939 818 B1

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
5 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
10 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
15 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
20 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
25 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
30 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
35 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
40 245 250 255

45

50

55

EP 0 939 818 B1

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 5 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 10 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 15 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 20 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 25 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 30 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460
 35 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 40 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 45 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 50 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 55 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620
 5 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640
 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655
 10 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 660 665 670
 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 675 680 685
 15 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 690 695 700
 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
 20 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
 25 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 755 760 765
 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 775 780
 30 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 785 790 795 800
 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 805 810 815
 35 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 820 825 830
 Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 835 840 845
 40 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860
 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys *
 865 870 875

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2862 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS

EP 0 939 818 B1

(B) LOCATION: 1..2862

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

5

10

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15	48
5	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30	96
10	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60	192
15	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
20	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
25	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
30	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
35	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
40	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
45	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
50	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
55	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816

EP 0 939 818 B1

	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
5	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
10	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
15	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
20	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
25	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
30	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
35	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
40	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
45	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
50	TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
55	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632

EP 0 939 818 B1

5	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680
	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
10	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
15	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
20	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920
	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
25	GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
30	GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
35	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
40	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
45	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
	ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
50	AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
55	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448

EP 0 939 818 B1

5	GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp	2496
	820 825 830	
	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val	2544
	835 840 845	
10	AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val	2592
	850 855 860	
	GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg	2640
15	865 870 875 880	
	CCT GGA CCG GAG ACG CTC TGC GGG GCT GAG CTG GTG GAT GCT CTT CAG Pro Gly Pro Glu Thr Leu Cys Gly Ala Glu Leu Val Asp Ala Leu Gln	2688
	885 890 895	
20	TTC GTG TGT GGA GAC AGG GGC TTT TAT TTC AAC AAG CCC ACA GGG TAT Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe Asn Lys Pro Thr Gly Tyr	2736
	900 905 910	
	GGC TCC AGC AGT CGG AGG GCG CCT CAG ACA GGT ATC GTG GAT GAG TGC Gly Ser Ser Ser Arg Arg Ala Pro Gln Thr Gly Ile Val Asp Glu Cys	2784
25	915 920 925	
	TGC TTC CGG AGC TGT GAT CTA AGG AGG CTG GAG ATG TAT TGC GCA CCC Cys Phe Arg Ser Cys Asp Leu Arg Arg Leu Glu Met Tyr Cys Ala Pro	2832
	930 935 940	
30	CTC AAG CCT GCC AAG TCA GCT GAA GCT TAG Leu Lys Pro Ala Lys Ser Ala Glu Ala *	2862
	945 950	

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 954 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

EP 0 939 818 B1

	Met	Gln	Phe	Val	Asn	Lys	Gln	Phe	Asn	Tyr	Lys	Asp	Pro	Val	Asn	Gly
	1				5				10						15	
5	Val	Asp	Ile	Ala	Tyr	Ile	Lys	Ile	Pro	Asn	Ala	Gly	Gln	Met	Gln	Pro
				20					25					30		
	Val	Lys	Ala	Phe	Lys	Ile	His	Asn	Lys	Ile	Trp	Val	Ile	Pro	Glu	Arg
			35					40					45			
10	Asp	Thr	Phe	Thr	Asn	Pro	Glu	Glu	Gly	Asp	Leu	Asn	Pro	Pro	Pro	Glu
	50						55					60				
	Ala	Lys	Gln	Val	Pro	Val	Ser	Tyr	Tyr	Asp	Ser	Thr	Tyr	Leu	Ser	Thr
	65					70				75						80
15	Asp	Asn	Glu	Lys	Asp	Asn	Tyr	Leu	Lys	Gly	Val	Thr	Lys	Leu	Phe	Glu
					85					90					95	
	Arg	Ile	Tyr	Ser	Thr	Asp	Leu	Gly	Arg	Met	Leu	Leu	Thr	Ser	Ile	Val
				100					105					110		
20																
25																
30																
35																
40																
45																
50																
55																

EP 0 939 818 B1

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125
 5 Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140
 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160
 10 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175
 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190
 15 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205
 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220
 20 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240
 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255
 25 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 30 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 35 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 40 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 45 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 50 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 55 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460

EP 0 939 818 B1

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 5 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 10 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 15 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 20 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605
 25 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620
 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640
 30 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655
 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 660 665 670
 35 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 675 680 685
 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 690 695 700
 40 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
 45 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 755 760 765
 50 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 775 780
 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 785 790 795 800
 55 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 805 810 815

EP 0 939 818 B1

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 820 825 830
 5 Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 835 840 845
 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860
 10 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg
 865 870 875 880
 Pro Gly Pro Glu Thr Leu Cys Gly Ala Glu Leu Val Asp Ala Leu Gln
 885 890 895
 15 Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe Asn Lys Pro Thr Gly Tyr
 900 905 910
 Gly Ser Ser Ser Arg Arg Ala Pro Gln Thr Gly Ile Val Asp Glu Cys
 915 920 925
 20 Cys Phe Arg Ser Cys Asp Leu Arg Arg Leu Glu Met Tyr Cys Ala Pro
 930 935 940
 Leu Lys Pro Ala Lys Ser Ala Glu Ala *
 945 950

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2724 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2724

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

EP 0 939 818 B1

	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT	48
	Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
	1 5 10 15	
5	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG	96
	Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
	20 25 30	
10	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC	144
	Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
	35 40 45	
	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA	192
	Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	
	50 55 60	
15	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA	240
	Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	
	65 70 75 80	
20	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG	288
	Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	
	85 90 95	
25		
30		
35		
40		
45		
50		
55		

	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
5	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
10	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
15	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
20	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
25	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
30	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
35	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
40	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
45	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
50	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
55	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104

EP 0 939 818 B1

5	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
10	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
15	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
20	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
25	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
30	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
35	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
40	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
45	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
50	TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
55	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632
60	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680
65	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
70	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
75	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
80	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
85	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920

EP 0 939 818 B1

	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
5	GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
10	GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
15	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
20	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
25	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
30	ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
	AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
35	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
	GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
40	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544
45	AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860	2592
	GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880	2640
50	CCT CAA TCT AAA GTT AAA AGA CAA ATA TTT TCA GGC TAT CAA TCT GAT Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 885 890 895	2688
55	ATT GAT ACA CAT AAT AGA ATT AAG GAT GAA TTA TGA Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu * 900 905	2724

EP 0 939 818 B1

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 908 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
290 295 300

EP 0 939 818 B1

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460
 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605
 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620
 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640
 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655

EP 0 939 818 B1

```

5      Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
      660                                     665                                     670
      Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
      675                                     680                                     685
10     Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
      690                                     695                                     700
      Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
      705                                     710                                     715                                     720
15     Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
      725                                     730                                     735
      Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
      740                                     745                                     750
      Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
      755                                     760                                     765
20     Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
      770                                     775                                     780
      Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
      785                                     790                                     795                                     800
25     Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
      805                                     810                                     815
      Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
      820                                     825                                     830
30     Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
      835                                     840                                     845
      Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
      850                                     855                                     860
35     Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg
      865                                     870                                     875                                     880
      Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp
      885                                     890                                     895
40     Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu
      900                                     905

```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3042 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION:1..3042

EP 0 939 818 B1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

5	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT	48
	Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
	1 5 10 15	

10

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	96
	20 25 30	
5	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	144
	35 40 45	
10	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Gly Asp Leu Asn Pro Pro Pro Glu	192
	50 55 60	
	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	240
	65 70 75 80	
15	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	288
	85 90 95	
20	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	336
	100 105 110	
	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	384
	115 120 125	
25	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr	432
	130 135 140	
30	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile	480
	145 150 155 160	
	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr	528
	165 170 175	
35	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	576
	180 185 190	
	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu	624
	195 200 205	
40	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu	672
	210 215 220	
45	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn	720
	225 230 235 240	
	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu	768
	245 250 255	
50	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys	816
	260 265 270	
55	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn	864
	275 280 285	

EP 0 939 818 B1

5	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
10	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
15	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
20	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
25	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
30	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
35	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
40	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
45	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
50	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
55	ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
60	AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
65	GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
70	GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
75	TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
80	AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632
85	CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680

5	CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
10	AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
15	TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
20	ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
25	ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920
30	ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
35	GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
40	GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
45	GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
50	GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
55	GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
60	AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
65	ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
70	ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
75	AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
80	TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
85	GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496

EP 0 939 818 B1

5	AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544
10	AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860	2592
15	GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCA GGC Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 865 870 875 880	2640
20	CTG AAT TCC CCG GGT GCA GCT CAT TAT GCG CAA CAC GAT GAA GCC GTA Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895	2688
25	GAC AAC AAA TTC AAC AAA GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA Asp Asn Lys Phe Asn Lys Glu Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910	2736
30	CAT TTA CCT AAC TTA AAC GAA GAA CAA CGA AAC GCC TTC ATC CAA AGT His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925	2784
35	TTA AAA GAT GAC CCA AGC CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940	2832
40	AAG CTA AAT GAT GCT CAG GCG CCG AAA GTA GAC AAC AAA TTC AAC AAA Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960	2880
45	GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975	2928
50	GAA GAA CAA CGA AAC GCC TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990	2976
55	CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA AAG CTA AAT GAT GCT CAG Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005	3024
60	GCG CCG AAA GTA GAC TAG Ala Pro Lys Val Asp * 1010	3042

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1014 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

EP 0 939 818 B1

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 1 5 10 15
 5 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
 20 25 30
 Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
 35 40 45

10

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
 50 55 60
 5 Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
 65 70 75 80
 Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
 85 90 95
 10 Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
 100 105 110
 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125
 15 Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140
 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160
 20 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175
 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190
 25 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205
 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220
 30 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240
 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255
 35 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 40 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 45 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 50 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 55 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400

EP 0 939 818 B1

5 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 10 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460
 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 15 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 20 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 25 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 30 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605
 35 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620
 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640
 40 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655
 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 660 665 670
 45 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 675 680 685
 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 690 695 700
 50 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
 55 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 755 760 765
 5 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 775 780
 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 785 790 795 800
 10 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 805 810 815
 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 820 825 830
 15 Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 835 840 845
 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860
 20 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly
 865 870 875 880
 Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val
 885 890 895
 25 Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu
 900 905 910
 His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser
 915 920 925
 30 Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys
 930 935 940
 Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys
 945 950 955 960
 35 Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn
 965 970 975
 Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser
 980 985 990
 40 Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln
 995 1000 1005
 Ala Pro Lys Val Asp *
 1010

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3509 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

EP 0 939 818 B1

(B) LOCATION:1..3509

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

5

10

15

20

25

30

35

40

45

50

55

5	ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	48
10	AAT AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
15	TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
20	AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60	192
25	ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80	240
30	ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
35	AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
40	ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
45	TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
50	CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160	480
55	TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	528
60	ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190	576
65	ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
70	AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220	672
75	GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240	720
80	GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asn Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
85	TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816

	GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
5	TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
10	AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320	960
	AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
15	AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
20	ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
	ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
25	AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400	1200
30	TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
	AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
35	AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445	1344
40	GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460	1392
	GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480	1440
45	TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488
	TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510	1536
50	GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525	1584
55	AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535 540	1632

5	ACA TTT CCT CTA GAT ATA AGA GAT ATA AGT TTA ACA TCT TCA TTT GAT Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 555 560	1680
10	GAT GCA TTA TTA TTT TCT AAC AAA GTT TAT TCA TTT TTT TCT ATG GAT Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575	1728
15	TAT ATT AAA ACT GCT AAT AAA GTG CTA GAA GCA GGA TTA TTT GCA GGT Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590	1776
20	TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AGC Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605	1824
25	AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT ATA Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 615 620	1872
30	GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GAA Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 630 635 640	1920
35	AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CCA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655	1968
40	GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATT Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670	2016
45	GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AAA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685	2064
50	AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TGG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700	2112
55	CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TAT Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720	2160
60	AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TAC Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735	2208
65	AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GAT Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750	2256
70	TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT ATA Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765	2304
75	GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATG Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 775 780	2352
80	AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAT Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800	2400
85	ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815	2448

EP 0 939 818 B1

5	TTG ATT GGA AGT GCA GAA TAT GAA AAA TCA AAA GTA AAT AAA TAC TTG Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830	2496
10	AAA ACC ATT ATG CCG TTT GAT CTT TCA ATA TAT ACC AAT GAT ACA ATA Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845	2544
15	CTA ATA GAA ATG TTT AAT AAA TAT AAT AGC GAA ATT TTA AAT AAT ATT Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855 860	2592
20	ATC TTA AAT TTA AGA TAT AAG GAT AAT AAT TTA ATA GAT TTA TCA GGA Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 865 870 875 880	2640
25	TAT GGG GCA AAG GTA GAG GTA TAT GAT GGA GTC CAG CTT AAT GAT AAA Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895	2688
30	AAT CAA TTT AAA TTA ACT AGT TCA GCA AAT AGT AAG ATT AGA GTG ACT Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910	2736
35	CAA AAT CAG AAT ATC ATA TTT AAT AGT GTG TTC CTT GAT TTT AGC GTT Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925	2784
40	AGC TTT TGG ATA AGA ATA CCT AAA TAT AAG AAT GAT GGT ATA CAA AAT Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940	2832
45	TAT ATT CAT AAT GAA TAT ACA ATA ATT AAT TGT ATG AAA AAT AAT TCG Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960	2880
50	GGC TGG AAA ATA TCT ATT AGG GGT AAT AGG ATA ATA TGG ACT TTA ATT Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975	2928
55	GAT ATA AAT GGA AAA ACC AAA TCG GTA TTT TTT GAA TAT AAC ATA AGA Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990	2976
60	GAA GAT ATA TCA GAG TAT ATA AAT AGA TGG TTT TTT GTA ACT ATT ACT Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005	3024
65	AAT AAT TTG AAT AAC GCT AAA ATT TAT ATT AAT GGT AAG CTA GAA TCA Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020	3072
70	AAT ACA GAT ATT AAA GAT ATA AGA GAA GTT ATT GCT AAT GGT GAA ATA Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040	3120
75	ATA TTT AAA TTA GAT GGT GAT ATA GAT AGA ACA CAA TTT ATT TGG ATG Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055	3168
80	AAA TAT TTC AGT ATT TTT AAT ACG GAA TTA AGT CAA TCA AAT ATT GAA Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070	3216
85	GAA AGA TAT AAA ATT CAA TCA TAT AGC GAA TAT TTA AAA GAT TTT TGG Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085	3264

EP 0 939 818 B1

5 GGA AAT CCT TTA ATG TAC AAT AAA GAA TAT TAT ATG TTT AAT GCG GGG 3312
 Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly
 1090 1095 1100

AAT AAA AAT TCA TAT ATT AAA CTA AAG AAA GAT TCA CCT GTA GGT GAA 3360
 Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu
 1105 1110 1115 1120

10 ATT TTA ACA CGT AGC AAA TAT AAT CAA AAT TCT AAA TAT ATA AAT TAT 3408
 Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr
 1125 1130 1135

AGA GAT TTA TAT ATT GGA GAA AAA TTT ATT ATA AGA AGA AAG TCA AAT 3456
 Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn
 1140 1145 1150

15 TCT CAA TCT ATA AAT GAT GAT ATA GTT AGA AAA GAA GAT TAT ATA TAT 3504
 Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr
 1155 1160 1165

20 CTA GA
 Leu 3509

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1169 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

EP 0 939 818 B1

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn
1 5 10 15
Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg
20 25 30
Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
35 40 45
Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly
50 55 60
Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn
65 70 75 80
Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
85 90 95
Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile
100 105 110
Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu
115 120 125
Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn
130 135 140
Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile
145 150 155 160
Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
165 170 175

30

35

40

45

50

55

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

5 Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu
195 200 205

10 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
210 215 220

Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
225 230 235 240

Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe
245 250 255

15 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe
260 265 270

Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile
275 280 285

20 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn
290 295 300

Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr
305 310 315 320

25 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly
325 330 335

Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu
340 345 350

30 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys
355 360 365

Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
370 375 380

35 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
385 390 395 400

Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile
405 410 415

40 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
420 425 430

Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
435 440 445

45 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser
450 455 460

Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn
465 470 475 480

50 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
485 490 495

Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
500 505 510

Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
515 520 525

EP 0 939 818 B1

5 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540
 Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp
 545 550 555 560
 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp
 565 570 575
 10 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly
 580 585 590
 Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser
 595 600 605
 15 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile
 610 615 620
 Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu
 625 630 635 640
 20 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro
 645 650 655
 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile
 660 665 670
 25 Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys
 675 680 685
 Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp
 690 695 700
 30 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr
 705 710 715 720
 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr
 725 730 735
 35 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp
 740 745 750
 Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile
 755 760 765
 40 Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met
 770 775 780
 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn
 785 790 795 800
 45 Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr
 805 810 815
 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu
 820 825 830
 50 Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile
 835 840 845
 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile
 850 855 860
 55 Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly
 865 870 875 880

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys
 885 890 895
 Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr
 900 905 910
 Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val
 915 920 925
 Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn
 930 935 940
 Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser
 945 950 955 960
 Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile
 965 970 975
 Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg
 980 985 990
 Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr
 995 1000 1005
 Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser
 1010 1015 1020
 Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile
 1025 1030 1035 1040
 Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met
 1045 1050 1055
 Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu
 1060 1065 1070
 Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp
 1075 1080 1085
 Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly
 1090 1095 1100
 Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu
 1105 1110 1115 1120
 Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr
 1125 1130 1135
 Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn
 1140 1145 1150
 Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr
 1155 1160 1165
 Leu

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2574 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

EP 0 939 818 B1

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

5

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

10

15

20

25

30

35

40

45

50

55

5	ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	48
10	AAT AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
15	TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
20	AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60	192
25	ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80	240
30	ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
35	AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
40	ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
45	TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
50	CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160	480
55	TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	528
60	ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190	576
65	ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
70	AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220	672
75	GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240	720

EP 0 939 818 B1

	GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
5	TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816
10	GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
	TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
15	AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320	960
20	AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
	AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
25	ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
30	ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
	AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400	1200
35	TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
	AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
40	AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445	1344
45	GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460	1392
	GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480	1440
50	TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488
55	TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510	1536

EP 0 939 818 B1

5	GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525	1584
10	AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535 540	1632
15	ACA TTT CCT CTA GAT ATA AGA GAT ATA AGT TTA ACA TCT TCA TTT GAT Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Met Asp 545 550 555 560	1680
20	GAT GCA TTA TTA TTT TCT AAC AAA GTT TAT TCA TTT TTT TCT ATG GAT Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575	1728
25	TAT ATT AAA ACT GCT AAT AAA GTG GTA GAA GCA GGA TTA TTT GCA GGT Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590	1776
30	TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AGC Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605	1824
35	AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT ATA Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 615 620	1872
40	GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GAA Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 630 635 640	1920
45	AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CCA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655	1968
50	GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATT Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670	2016
55	GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AAA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685	2064
60	AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TGG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700	2112
65	CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TAT Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720	2160
70	AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TAC Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735	2208
75	AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GAT Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750	2256
80	TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT ATA Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765	2304
85	GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATG Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 775 780	2352

EP 0 939 818 B1

5 AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAT 2400
 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn
 785 790 795 800

ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT 2448
 Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr
 805 810 815

10 TTG ATT GGA AGT GCA GAA TAT GAA AAA TCA AAA GTA AAT AAA TAC TTG 2496
 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu
 820 825 830

AAA ACC ATT ATG CCG TTT GAT CTT TCA ATA TAT ACC AAT GAT ACA ATA 2544
 Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile
 835 840 845

15 CTA ATA GAA ATG TTT AAT AAA TAT AAT AGC 2574
 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser
 850 855

20 (2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 858 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

35

40

45

50

55

EP 0 939 818 B1

1 Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn
 5 5 10 15
 20 Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg
 25 30
 35 Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
 40 45
 50 Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly
 55 60
 65 Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn
 70 75 80
 85 Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
 90 95
 100 Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile
 105 110
 115 Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu
 120 125
 130 Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn
 135 140
 145 Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile
 150 155 160
 165 Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
 170 175
 180 Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
 185 190

EP 0 939 818 B1

5 Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu
 195 200 205
 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
 210 215 220
 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
 225 230 235 240
 10 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe
 245 250 255
 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe
 260 265 270
 15 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile
 275 280 285
 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn
 290 295 300
 20 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr
 305 310 315 320
 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly
 325 330 335
 25 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu
 340 345 350
 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys
 355 360 365
 30 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
 370 375 380
 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
 385 390 395 400
 35 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile
 405 410 415
 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
 420 425 430
 40 Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
 435 440 445
 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser
 450 455 460
 45 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn
 465 470 475 480
 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
 485 490 495
 50 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
 500 505 510
 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
 515 520 525
 55 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp
 545 550 555 560
 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp
 565 570 575
 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly
 580 585 590
 Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser
 595 600 605
 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile
 610 615 620
 Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu
 625 630 635 640
 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro
 645 650 655
 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile
 660 665 670
 Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys
 675 680 685
 Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp
 690 695 700
 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr
 705 710 715 720
 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr
 725 730 735
 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp
 740 745 750
 Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile
 755 760 765
 Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met
 770 775 780
 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn
 785 790 795 800
 Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr
 805 810 815
 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu
 820 825 830
 Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile
 835 840 845
 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser
 850 855

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1644 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

EP 0 939 818 B1

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1..1644

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23;

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

	ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	48
5	AAT AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
10	TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
	AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60	192
15	ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80	240
20	ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
	AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
25	ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
30	TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
	CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160	480
35	TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	528
40	ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190	576
	ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
45	AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220	672

	GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240	720
5	GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
10	TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816
	GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
15	TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
20	AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320	960
	AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
25	AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
30	ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
	ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
35	AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400	1200
40	TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
	AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
45	AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445	1344
50	GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460	1392
	GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480	1440
55	TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488

EP 0 939 818 B1

5 TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT 1536
 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
 500 505 510

10 GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA 1584
 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
 515 520 525

15 AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG 1632
 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540

20 ACA TTT CCT CTA 1644
 Thr Phe Pro Leu
 545

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 548 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

EP 0 939 818 B1

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn
 1 5 10 15
 5 Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg
 20 25 30
 Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
 35 40 45
 10 Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly
 50 55 60
 Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn
 65 70 75 80
 15 Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
 85 90 95
 Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile
 100 105 110
 20 Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu
 115 120 125
 Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn
 130 135 140
 25 Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile
 145 150 155 160
 Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
 165 170 175
 30 Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
 180 185 190
 Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu
 195 200 205

35

40

45

50

55

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
 210 215 220
 5 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
 225 230 235 240
 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe
 245 250 255
 10 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe
 260 265 270
 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile
 275 280 285
 15 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn
 290 295 300
 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr
 305 310 315 320
 20 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly
 325 330 335
 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu
 340 345 350
 25 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys
 355 360 365
 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
 370 375 380
 30 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
 385 390 395 400
 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile
 405 410 415
 35 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
 420 425 430
 Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
 435 440 445
 40 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser
 450 455 460
 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn
 465 470 475 480
 45 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
 485 490 495
 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
 500 505 510
 50 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
 515 520 525
 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540
 55 Thr Phe Pro Leu
 545

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2616 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

1	ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT	48
5	Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
10	1 5 10 15	
20	GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG	96
25	Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
30	20 25 30	
35	GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC	144
40	Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
45	35 40 45	
50	GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA	192
55	Asp Thr Phe Thr Asn Pro Glu Gly Asp Leu Asn Pro Pro Pro Glu	
60	50 55 60	
65	GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA	240
70	Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	
75	65 70 75 80	
80	GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG	288
85	Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	
90	85 90 95	
95	CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC	336
100	Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	
105	100 105 110	
110	CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG	384
115	Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	
120	115 120 125	
125	GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC	432
130	Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr	
135	130 135 140	
140	AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT	480
145	Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile	
150	145 150 155 160	
155	ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG	528
160	Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr	
165	165 170 175	
170	CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC	576
175	Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	
180	180 185 190	

EP 0 939 818 B1

	ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu	624
	195 200 205	
5	GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu	672
	210 215 220	
10	CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn	720
	225 230 235 240	
	CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu	768
	245 250 255	
15	GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys	816
	260 265 270	
20	TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn	864
	275 280 285	
	AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val	912
	290 295 300	
25	GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys	960
	305 310 315 320	
	TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu	1008
	325 330 335	
30	AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp	1056
	340 345 350	
35	AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn	1104
	355 360 365	
	TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr	1152
	370 375 380	
40	ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn	1200
	385 390 395 400	
45	TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu	1248
	405 410 415	
	AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg	1296
	420 425 430	
50	GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys	1344
	435 440 445	
55	GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe	1392
	450 455 460	

EP 0 939 818 B1

	AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480	1440
5	ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT AGT TTA Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	1488
10	GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510	1536
	GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
15	GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
20	TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560	1680
	CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
25	TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
	AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
30	CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
35	GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640	1920
	TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT GTA GGT GCT TTA Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655	1968
40	ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
	ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
45	GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
50	AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720	2160
	GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735	2208

5	GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750	2256
10	CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT GAT GAT Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765	2304
15	TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780	2352
20	AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT TCT ATG Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800	2400
25	ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT CTT AAA Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815	2448
30	GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA ATT GGT Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830	2496
35	CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT ACA GAT Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845	2544
40	ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA TTA TCT Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860	2592
45	ACA TTT ACT GAA TAT ATT AAG TAA Thr Phe Thr Glu Tyr Ile Lys * 865 870	2616

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 872 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

EP 0 939 818 B1

1 Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 5 5 10 15
 20 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
 25 30
 35 Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
 40 45
 50 Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
 55 60
 65 Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
 70 75 80
 85 Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
 90 95
 100 Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
 105 110

20

25

30

35

40

45

50

55

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125
 5 Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140
 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160
 10 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175
 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190
 15 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205
 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220
 20 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240
 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255
 25 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 30 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 35 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 40 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 45 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 50 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 55 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
 450 455 460

EP 0 939 818 B1

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480
 5 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 10 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525
 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540
 15 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560
 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575
 20 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590
 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605
 25 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620
 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640
 30 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655
 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670
 35 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685
 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700
 40 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720
 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735
 45 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750
 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765
 50 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
 770 775 780
 Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
 785 790 795 800
 55 Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
 805 810 815

EP 0 939 818 B1

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
 820 825 830
 5 Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
 835 840 845
 Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
 850 855 860
 10 Thr Phe Thr Glu Tyr Ile Lys *
 865 870

(2) INFORMATION FOR SEQ ID NO: 27:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2574 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

30

35

40

45

50

55

	ATGCCGGTTA CCATCAACAA CTTCAACTAC AACGACCCGA TCGACAACAA CAACATCATC	60
	ATGATGGAAC CGCCGTTTCGC ACGTGGTACC GGTCGTTACT ACAAGGCTTT CAAGATCACC	120
5	GACCGTATCT GGATCATCCC GGAACGTTAC ACCTTCGGTT ACAAACCTGA GGACTTCAAC	180
	AAGAGTAGCG GGATTTTCAA TCGTGACGTC TGCGAGTACT ATGATCCAGA TTATCTGAAT	240
	ACCAACGATA AGAAGAACAT ATTCCTTCAG ACTATGATCA AGTTATTTAA TAGAATCAAA	300
10	TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGG	360
	GATAGACGTG TTCCACTCGA AGAGTTTAAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	420
	TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	480
15	TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAATCAT	540
	TTTGCATCAA GGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTGGCCC AGAATATGTA	600
	AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	660
20	TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTTATAC ATGTTTTACA TGGATTATAT	720
	GGCAITAAAG TAGATGATTT ACCAATTGTA CCAATGAAA AAAAATTTTT TATGCAATCT	780
	ACAGATGCTA TACAGGCAGA AGAACTATAT ACATTTGGAG GACAAGATCC CAGCATCATA	840
25	ACTCCTTCTA CGGATAAAAG TATCTATGAT AAAGTTTTGC AAAATTTTAG AGGGATAGTT	900
	GATAGACTTA ACAAGGTTTT AGTTTGCATA TCAGATCCTA ACATTAATAT TAATATATAT	960
	AAAAATAAAT TTAAAGATAA ATATAAATTC GTTGAAGATT CTGAGGGAAA ATATAGTATA	1020
30	GATGTAGAAA GTTTTGATAA ATTATATAAA AGCTTAATGT TTGGTTTTAC AGAAACTAAT	1080
	ATAGCAGAAA ATTATAAAAT AAAAAGTAGA GCTTCTTATT TTAGTGATTC CTTACCACCA	1140
	GTAAAAATAA AAAATTTATT AGATAATGAA ATCTATACTA TAGAGGAAGG GTTTAATATA	1200

35

40

45

50

55

	TCTGATAAAG ATATGGAAAA AGAATATAGA GGTCAGAATA AAGCTATAAA TAAACAAGCT	1260
	TATGAAGAAA TTAGCAAGGA GCATTGGCT GTATATAAGA TACAAATGTG TAAAAGTGTT	1320
5	AAAGCTCCAG GAATATGTAT TGATGTTGAT AATGAAGATT TGTTCTTTAT AGCTGATAAA	1380
	AATAGTTTTT CAGATGATTT ATCTAAAAAC GAAAGAATAG AATATAATAC ACAGAGTAAT	1440
	TATATAGAAA ATGACTTCCC TATAAATGAA TTAATTTTAG ATACTGATTT AATAAGTAAA	1500
10	ATAGAATTAC CAAGTGAAAA TACAGAATCA CTTACTGATT TTAATGTAGA TGTTCCAGTA	1560
	TATGAAAAAC AACCCGCTAT AAAAAAATT TTTACAGATG AAAATACCAT CTTTCAATAT	1620
	TTTACTCTC AGACATTTCC TCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
15	GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAACT	1740
	GCTAATAAG TGGTAGAAGC AGGATTATTT GCAGGTGGG TGAACAGAT AGTAAATGAT	1800
	TTTGTAAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
20	GTTCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
	AATGCTTTTG AGATTGCAGG AGCCAGTATT CTTACTAGAAT TTATACCAGA ACTTTTAATA	1980
	CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
25	ACAATAGATA ATGCTTTAAC TAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
	GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAGA GGAATGTAT	2160
	AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
30	TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAACTT	2280
	AATGAGGGTA TTAACCAAGC TATAGATAAT ATAAATAATT TTATAAATGG ATGTTCTGTA	2340
	TCATATTTAA TGAAAAAAT GATTCCATTA GCTGTAGAAA AATTACTAGA CTTTGATAAT	2400
35	ACTCTCAAAA AAAATTTGTT AAATTATATA GATGAAAATA AATTATATTT GATTGGAAGT	2460
	GCAGAATATG AAAAATCAAA AGTAAATAAA TACTTGAAAA CCATTATGCC GTTTGATCTT	2520
	TCAATATATA CCAATGATAC AATACTAATA GAAATGTTTA ATAAATATAA TAGC	2574

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2574 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

EP 0 939 818 B1

	ATGCCAGTTA CAATAAATAA TTTTAATTAT AATGATCCTA TTGATAATAA TAATATTATT	60
	ATGATGGAGC CTCCATTTCG GAGAGGTACG GGGAGATATT ATAAAGCTTT TAAAATCACA	120
5	GATCGTATTT GGATAATACC GGAAAGATAT ACTTTTGGAT ATAAACCTGA GGATTTTAAT	180
	AAAAGTTCCG GTATTTTAA TAGAGATGTT TGTGAATATT ATGATCCAGA TTACTTAAAT	240

10

15

20

25

30

35

40

45

50

55

EP 0 939 818 B1

	ACTAATGATA AAAAGAATAT ATTTTACAA ACAATGATCA AGTTATTTAA TAGAATCAAA	300
	TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	360
5	GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	420
	TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAA TTTAATAATA	480
	TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAATCAT	540
10	TTTGCATCAA GGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	600
	AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCACAAGTA TATTTAATAG ACGTGGATAT	660
	TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACGTCTCCA CGGTCTCTAC	720
15	GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAATTCTT CATGCAGAGC	780
	ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	840
	ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCCTCC AGAACTTCG TGGTATCGTC	900
20	GACCGTCTCA ACAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	960
	AAAAACAAAT TCAAAGACAA ATACAAATTC GTCGAAGACA GCGAAGGTAA ATACAGCATC	1020
	GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1080
25	ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1140
	GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTTCAACATC	1200
	AGCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACA AAGCAATCAA CAAACAAGCT	1260
30	TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAA TCCAGATGTG CAAAAGCGTC	1320
	AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1380
	AACAGCTTCA GCGACGACCT CAGCAAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1440
35	TACATCGAAA ACGACTTCCC GATCAACGAA CTCATCCTCG ACACCGACCT CATCAGCAA	1500
	ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
	TACGAAAAAC AGCCGGCAAT CAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
40	CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
	GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAACT	1740
	GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
45	TTTGTAAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
	GTTCCCTTATA TAGGATTAGC TTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
	AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
50	CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATAITGACA ATAAAAATAA AATTATTAAA	2040
	ACAATAGATA ATGCTTTAAC TAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
	GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAGA GGAATGTAT	2160
55	AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
	TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAAACTT	2280

AATGAGGGTA TTAACCAAGC TATAGATAAT ATAAATAATT TTATAAATGG ATGTTCTGTA 2340
 TCATATTTTAA TGAAAAAAT GATTCCATTA GCTGTAGAAA AATTACTAGA CTTTGATAAT 2400
 5 ACTCTCAAAA AAAATTTGTT AAATTATATA GATGAAAATA AATTATATTT GATTGGAAGT 2460
 GCAGAATATG AAAAATCAAA AGTAAATAAA TACTTGAAAA CCATTATGCC GTTTGATCTT 2520
 10 TCAATATATA CCAATGATAC AATACTAATA GAAATGTTTA ATAAATATAA TAGC 2574

15

Claims

1. A polypeptide composing first and second domains, wherein
 said first domain is a clostridial neurotoxin light chain or a fragment or variant of said clostridial neurotoxin
 light chain, wherein said first domain is capable of cleaving one or more vesicle or plasma-membrane associated
 proteins essential to exocytosis,
 said second domain is a clostridial neurotoxin heavy chain H_N portion or a fragment or variant of said clostridial
 neurotoxin heavy chain H_N portion, wherein said second domain is capable of (i) translocating the polypeptide into
 a cell or (ii) increasing the solubility of the polypeptide compared to the solubility of the first domain on its own or
 25 (iii) both translocating the polypeptide into a cell and increasing the solubility of the polypeptide compared to the
 solubility of the first domain on its own,
 wherein said second domain lacks an intact portion designated H_C of a clostridial neurotoxin heavy chain, and
 said polypeptide is a single chain polypeptide.
2. A polypeptide according to Claim 1 wherein the clostridial toxin is a botulinum toxin.
3. A polypeptide according to Claim 1 or Claim 2 wherein the first domain exhibits endopeptidase activity specific for
 a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
4. A polypeptide according to Claim 1 wherein the clostridial toxin heavy chain H_N portion is from a botulinum toxin.
5. A polypeptide according to any of Claims 1-4 further comprising a third domain that binds of the polypeptide to a
 cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that
 bind to a cell.
6. A polypeptide according to Claim 5 wherein said third domain is for binding the polypeptide to an immunoglobulin.
7. A polypeptide according to Claim 6 wherein said third domain is a tandem repeat synthetic IgG binding domain
 derived from domain β of Staphylococcal protein A.
8. A polypeptide according to Claim 5 wherein said third domain comprises an amino acid sequence that binds to a
 cell surface receptor.
9. A polypeptide according to Claim 8 wherein said third domain is insulin-like growth factor-1 (IGF-1).
10. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant
 of a botulinum toxin light chain and a portion designated H_N of a botulinum toxin heavy chain.
11. A polypeptide according to Claim 10 wherein one or both of (a) the toxin light chain or fragment or variant of toxin
 light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
12. A polypeptide according to Claim 11 wherein the botulinum toxin type A light chain variant has at residue 2 a
 glutamate, at residue 26 a lysine, and at residue 27 a tyrosine.

13. A polypeptide according to Claim 10 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
14. A polypeptide according to any of Claims 1-9 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
15. A polypeptide according to Claim 14 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
16. A polypeptide according to Claim 11 or 12 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
17. A polypeptide according to Claim 16 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
18. A polypeptide according to Claim 16 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
19. A polypeptide according to Claim 13 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
20. A polypeptide according to Claim 19 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
21. A polypeptide according to any of Claims 10-20 lacking a portion designated H_C of a botulinum toxin heavy chain.
22. A polypeptide according to any of Claims 1-4, wherein the second domain is not capable of binding to cell surface receptors.
23. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
24. A polypeptide according to Claim 23 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
25. A polypeptide according to Claim 23 or 24 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
26. A polypeptide according to Claim 23, 24 or 25 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
27. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-26 with (b) a second polypeptide being a polypeptide or oligopeptide that binds to an affinity matrix so as to enable purification of the fusion protein using said matrix.
28. A fusion protein according to Claim 27 wherein said second polypeptide binds to a chromatography column.
29. A fusion protein according to Claim 28 wherein said second chromatography column is an affinity matrix of glutathione sepharose.
30. A fusion protein according to Claim 27, 28 or 29 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
31. A composition comprising a polypeptide according to any preceding claim, said composition being non-toxic *in vivo*.

EP 0 939 818 B1

32. A composition according to Claim 31 or a polypeptide according to any of Claims 1-26 or a fusion protein according to Claim 27, 28, 29 or 30 for use as a positive control in a toxin assay.
33. A composition according to Claim 31 or a polypeptide according to any of Claims 1-26 or a fusion protein according to Claim 27, 28, 29 or 30 for use as a vaccine against clostridial toxin.
34. A composition according to Claim 31 or a polypeptide according to any of Claims 1-26 or a fusion protein according to Claim 27, 28, 29, or 30 for *in vivo* use.
35. A pharmaceutical composition comprising a composition according to Claim 31, a polypeptide according to any of claims 1-26 or a fusion protein according to Claim 27, 28, 29 or 30, in combination with a pharmaceutically acceptable carrier.
36. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-30.
37. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 36 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.
38. A nucleic acid according to Claim 36 or 37 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H_N domain.
39. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 36 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
40. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 36 or 39 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H_N domain.
41. A nucleic acid according to any of Claims 36-40 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
42. A nucleic acid according to Claim 41 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-30 so as to introduce said cleavage site.
43. A DNA according to any of Claims 36-42.
44. A DNA selected from SEQ ID No:s 1, 3, 5, 9, 11, 13, 15, 17, 19, 21, 27 and 28.
45. A method of manufacture of a polypeptide according to any of Claims 1-26 comprising expressing in a host cell a nucleic acid according to any of Claims 36-44 and recovering the polypeptide.
46. A method of manufacture of a polypeptide according to any of Claims 1-26 comprising expressing in a host cell a nucleic acid encoding a fusion protein according to Claim 27, 28, 29 or 30, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.
47. A method of manufacture according to Claims 45 or 46 in which the nucleic acid is DNA.
48. A cell expressing a polypeptide or fusion protein according to any of Claims 1-30.

Patentansprüche

1. Polypeptid, umfassend erste und zweite Domänen,
worin die erste Domäne die leichte Kette eines Clostridienneurotoxins oder ein Fragment oder eine Variante der leichten Kette des Clostridienneurotoxins ist, worin die erste Domäne einen oder mehrere Vesikel- oder Plasmamembranassoziierte Proteine, die für die Exocytose wesentlich sind, spalten kann,
worin die zweite Domäne ein H_N-Teil der schweren Kette eines Clostridienneurotoxins oder ein Fragment oder eine Variante des H_N-Teils der schweren Kette des Clostridienneurotoxins ist, worin die zweite Domäne (i)

das Polypeptid in eine Zelle translozieren kann oder (ii) die Löslichkeit des Polypeptids im Vergleich zur Löslichkeit der ersten Domäne alleine erhöhen kann oder (iii) sowohl das Polypeptid in eine Zelle translozieren als auch die Löslichkeit des Polypeptids im Vergleich zur Löslichkeit der ersten Domäne alleine erhöhen kann, worin der zweiten Domäne ein zweiter, als H_C-bezeichneter intakter Teil der schweren Kette des Clostridienneurotoxins fehlt, und worin das Polypeptid ein Einzelkettenpolypeptid ist.

2. Polypeptid nach Anspruch 1, worin das Clostridientoxin ein Botulinustoxin ist.
3. Polypeptid nach Anspruch 1 oder Anspruch 2, worin die erste Domäne eine Endopeptidaseaktivität aufweist, die spezifisch für ein Substrat ist, ausgewählt unter einem oder mehreren von SNAP-25, Synaptobrevin/VAMP und Syntaxin.
4. Polypeptid nach Anspruch 1, worin der H_N-Teil der schweren Kette des Clostridienneurotoxins von einem Botulinustoxin ist.
5. Polypeptid nach einem der Ansprüche 1 bis 4, zusätzlich umfassend eine dritte Domäne, die von dem Polypeptid an eine Zelle bindet, und zwar indem die dritte Domäne direkt an eine Zelle bindet oder indem die dritte Domäne an einen Ligand oder an Liganden bindet, der bzw. die an eine Zelle bindet bzw. binden.
6. Polypeptid nach Anspruch 5, worin die dritte Domäne für die Bindung des Polypeptids an ein Immunglobulin ist.
7. Polypeptid nach Anspruch 6, worin die dritte Domäne eine synthetische Tandemrepeat- IgG-Bindungsdomäne ist, abgeleitet von der Domäne β des Staphylococcenproteins A.
8. Polypeptid nach Anspruch 5, worin die dritte Domäne eine Aminosäuresequenz umfasst, die an einen Zelloberflächenrezeptor bindet.
9. Polypeptid nach Anspruch 8, worin die dritte Domäne der insulinähnliche Wachstumsfaktor-1 (IFG-1 = insulin-like growth factor-1) ist.
10. Polypeptid nach einem der vorstehenden Ansprüche, umfassend eine leichte Kette eines Botulinustoxins oder ein Fragment oder eine Variante einer leichten Kette eines Botulinustoxins und einen als H_N-bezeichneten Teil der schweren Kette eines Botulinustoxins.
11. Polypeptid nach Anspruch 10, worin eines oder beide von (a) der leichten Kette des Toxins oder des Fragments oder der Variante der leichten Kette des Toxins und (b) des Teils der schweren Kette des Toxins vom Botulinustoxin Typ A sind.
12. Polypeptid nach Anspruch 11, worin die Variante der leichten Kette des Botulinustoxins Typ A an Rest 2 ein Glutamat, an Rest 26 ein Lysin und an Rest 27 ein Tyrosin aufweist.
13. Polypeptid nach Anspruch 10, worin eines oder beide von (a) der leichten Kette des Toxins oder des Fragments oder der Variante der leichten Kette des Toxins und (b) des Teils der schweren Kette des Toxins vom Botulinustoxin Typ B sind.
14. Polypeptid nach einem der Ansprüche 1-9, umfassend eine leichte Kette eines Botulinustoxins oder ein Fragment oder eine Variante einer leichten Kette eines Botulinustoxins und mindestens 100 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins.
15. Polypeptid nach Anspruch 14, umfassend eine leichte Kette eines Botulinustoxins Typ B oder ein Fragment oder eine Variante davon und 107 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins Typ B.
16. Polypeptid nach Anspruch 11 oder 12, umfassend mindestens 423 der N-terminalen Aminosäuren der schweren Kette des Botulinustoxins Typ A.
17. Polypeptid gemäß Anspruch 16, umfassend eine leichte Kette eines Botulinustoxins Typ A und 423 N-terminale Aminosäuren der schweren Kette eines Botulinustoxins Typ A.

18. Polypeptid gemäß Anspruch 16, umfassend eine Variante der leichten Kette des Botulinustoxins Typ A, worin Rest 2 ein Glutamat, Rest 26 ein Lysin und Rest 27 ein Tyrosin ist, und 423 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins Typ A.
- 5 19. Polypeptid nach Anspruch 13, umfassend mindestens 417 der N-terminalen Aminosäuren der schweren Kette des Botulinustoxins Typ B.
20. Polypeptid nach Anspruch 19, umfassend eine leichte Kette eines Botulinustoxins Typ B und 417 N-terminale Aminosäuren einer schweren Kette eines Botulinustoxins Typ B.
- 10 21. Polypeptid nach einem der Ansprüche 10-20, dem ein als H_C-bezeichneter Teil einer schweren Kette eines Botulinustoxins fehlt.
- 15 22. Polypeptid nach einem der Ansprüche 1-4, worin die zweite Domäne nicht an Zelloberflächenrezeptoren binden kann.
- 20 23. Polypeptid nach einem der vorstehenden Ansprüche, umfassend eine Variante eines Clostridientoxins und zusätzlich umfassend eine Stelle zur Spaltung durch ein proteolytisches Enzym, welche Spaltstelle im nativen Toxin nicht vorhanden ist.
- 25 24. Polypeptid nach Anspruch 23, umfassend eine Variante der leichten Kette eines Clostridientoxins und zusätzlich umfassend eine Stelle zur Spaltung durch ein proteolytisches Enzym, worin die Spaltstelle in der leichten Kette des nativen Toxins nicht vorhanden ist.
- 30 25. Polypeptid nach Anspruch 23 oder 24, umfassend eine Variante des H_N-Teils einer schweren Kette eines Clostridientoxins und zusätzlich umfassend eine Stelle zur Spaltung durch ein proteolytisches Enzym, welche Spaltstelle in dem nativen H_N-Teil der schweren Kette des Toxins nicht vorhanden ist.
- 35 26. Polypeptid nach Anspruch 23, 24 oder 25, erhältlich durch Modifikation einer das Polypeptid kodierenden DNS, um ein oder mehrere Nukleotide einzuführen, welche für die Spaltstelle kodieren.
27. Fusionsprotein, umfassend eine Fusion von (a) einem Polypeptid nach einem der Ansprüche 1-26 mit (b) einem zweiten Polypeptid, bei dem es sich um ein Polypeptid oder Oligopeptid handelt, das an eine Affinitätsmatrix bindet, um die Reinigung des Fusionsproteins unter Verwendung der Matrix zu ermöglichen.
- 40 28. Fusionsprotein gemäß Anspruch 27, worin das zweite Polypeptid an eine Chromatographiesäule bindet.
29. Fusionsprotein nach Anspruch 28, worin die zweite Chromatographiesäule eine Affinitätsmatrix von Glutathionsepharose ist.
- 30 30. Fusionsprotein nach Anspruch 27, 28 oder 29, worin eine spezifische Proteasespaltstelle zwischen dem ersten und zweiten Polypeptid eingebaut ist, worin die Proteasestelle die proteolytische Trennung des ersten und zweiten Polypeptids ermöglicht.
- 45 31. Zusammensetzung, umfassend ein Polypeptid nach einem der vorstehenden Ansprüche, worin die Zusammensetzung *in vivo* nicht toxisch ist.
- 50 32. Zusammensetzung nach Anspruch 31 oder ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 zur Verwendung als Positivkontrolle in einem Toxintest.
33. Zusammensetzung nach Anspruch 31 oder ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 zur Verwendung als ein Vaccin gegen ein Clostridientoxin.
- 55 34. Zusammensetzung gemäß Anspruch 31 oder ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 für die *in vivo*-Verwendung.
35. Pharmazeutische Zusammensetzung, umfassend eine Zusammensetzung nach Anspruch 31, ein Polypeptid nach einem der Ansprüche 1-26 oder ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 in Kombination mit einem

pharmazeutisch annehmbaren Träger.

36. Nukleinsäure, kodierend ein Polypeptid oder ein Fusionsprotein nach einem der Ansprüche 1-30.

5 37. Nukleinsäure nach Anspruch 36, kodierend ein Polypeptid oder ein Fusionsprotein und umfassend Nukleotide, kodierend die Reste 1-448 einer leichten Kette eines Botulinustoxins Typ A.

38. Nukleinsäure gemäß Anspruch 36 oder 37, umfassend Nukleotide, kodierend die Reste 1-423 einer H_N-Domäne der schweren Kette eines Botulinustoxins Typ A.

10 39. Nukleinsäure nach Anspruch 36, kodierend ein Polypeptid oder ein Fusionsprotein und umfassend Nukleotide, kodierend die Reste 1-470 der leichten Kette eines Botulinustoxins Typ B.

15 40. Nukleinsäure nach Anspruch 36 oder 39, kodierend ein Polypeptid oder ein Fusionsprotein und umfassend Nukleotide, kodierend die Reste 1-417 einer H_N-Domäne der schweren Kette des Botulinustoxins Typ B.

41. Nukleinsäure nach einem der Ansprüche 36-40, umfassend Nukleotide, kodierend eine Restriktionsendonucleasespaltstelle, die in der nativen Sequenz des Clostridientoxins nicht vorhanden ist.

20 42. Nukleinsäure nach Anspruch 41, erhältlich durch Modifikation eines Nukleotids, kodierend ein Polypeptid oder Fusionsprotein nach einem der Ansprüche 1-30, um die Spaltstelle einzuführen.

43. DNS nach einem der Ansprüche 36-42.

25 44. DNS, ausgewählt unter den SEQ IDs der Nummern 1, 3, 5, 9, 11, 13, 15, 17, 19, 21, 27 und 28.

45. Verfahren zur Herstellung eines Polypeptids nach einem der Ansprüche 1-26, umfassend das Expressieren einer Nukleinsäure nach einem der Ansprüche 36-44 in einer Wirtszelle und die Gewinnung des Polypeptids.

30 46. Verfahren der Herstellung eines Polypeptids nach einem der Ansprüche 1-26, umfassend das Expressieren einer Nukleinsäure, kodierend ein Fusionsprotein nach Anspruch 27, 28, 29 oder 30 in einer Wirtszelle, das Reinigen des Fusionsproteins mittels Elution des Fusionsproteins durch eine Affinitätsmatrix, angepasst an das Zurückhalten des Fusionsproteins, und Eluieren eines Liganden durch die Matrix, der an das Verdrängen des Fusionsproteins angepasst ist, und das Gewinnen des Fusionsproteins.

35 47. Verfahren zur Herstellung nach Anspruch 45 oder 46, worin die Nukleinsäure DNS ist.

48. Zelle, exprimierend ein Polypeptid oder Fusionsprotein nach einem der Ansprüche 1-30.

Revendications

1. Polypeptide comprenant un premier et un second domaines, dans lequel
 45 ledit premier domaine est une chaîne légère de neurotoxine clostridienne ou un fragment ou un variant de ladite chaîne légère de neurotoxine clostridienne, ledit premier domaine étant capable de cliver une ou plusieurs protéines associées à une membrane vésiculaire ou plasmique et essentielles pour l'exocytose,
 ledit deuxième domaine est une portion H_N de chaîne lourde de neurotoxine clostridienne ou un fragment ou un variant de la dite portion H_N de chaîne lourde de neurotoxine clostridienne, ledit deuxième domaine étant
 50 capable de (i) transloquer le polypeptide dans une cellule ou (ii) accroître la solubilité du polypeptide en comparaison avec la solubilité du premier domaine à lui seul ou (iii) à la fois transloquer le polypeptide dans une cellule et accroître la solubilité du polypeptide en comparaison avec la solubilité du premier domaine à lui seul,
 une portion intacte désignée H_C d'une chaîne lourde de neurotoxine clostridienne étant absente dudit deuxième domaine, et
 55 ledit polypeptide est un polypeptide à chaîne unique.

2. Polypeptide selon la revendication 1, dans lequel la toxine clostridienne est une toxine botulinique.

3. Polypeptide selon la revendication 1 ou la revendication 2, dans lequel le premier domaine possède une activité

endopeptidase spécifique d'un substrat choisi parmi un ou plusieurs de SNAP-25, synaptobrevine/VAMP et syntaxine.

- 5 4. Polypeptide selon la revendication 1, dans lequel la portion H_N de chaîne lourde de toxine clostridienne provient d'une toxine botulinique.
- 10 5. Polypeptide selon l'une quelconque des revendications 1 à 4 comprenant en outre un troisième domaine liant le polypeptide à une cellule, en liant le troisième domaine directement à une cellule ou en liant le troisième domaine à un ligand ou à des ligands qui se lient à une cellule.
- 15 6. Polypeptide selon la revendication 5, dans lequel ledit troisième domaine est destiné à lier le polypeptide à une immunoglobuline.
- 20 7. Polypeptide selon la revendication 6, dans lequel ledit troisième domaine est un domaine synthétique de liaison à l'IgG à séquences répétées en tandem dérivé du domaine bêta de la protéine A staphylococcique.
- 25 8. Polypeptide selon la revendication 5, dans lequel ledit troisième domaine comprend une séquence d'acides aminés qui se lie à un récepteur de surface cellulaire.
- 30 9. Polypeptide selon la revendication 8, dans lequel ledit troisième domaine est le facteur de croissance insulino-mimétique 1 (IGF-1).
- 35 10. Polypeptide selon l'une quelconque des revendications précédentes, comprenant une chaîne légère de toxine botulinique ou un fragment ou un variant d'une chaîne légère de toxine botulinique et une portion désignée H_N d'une chaîne lourde de toxine botulinique.
- 40 11. Polypeptide selon la revendication 10, dans lequel l'un ou les deux parmi (a) la chaîne légère de toxine ou le fragment ou le variant de chaîne légère de toxine et (b) la portion de la chaîne lourde de toxine, sont ceux d'une toxine botulinique de type A.
- 45 12. Polypeptide selon la revendication 11, dans lequel le variant de chaîne légère de toxine botulinique de type A possède un glutamate au résidu 2, une lysine au résidu 26, et une tyrosine au résidu 27.
- 50 13. Polypeptide selon la revendication 10, dans lequel l'un ou les deux parmi (a) la chaîne légère de toxine ou le fragment ou le variant de chaîne légère de toxine et (b) la portion de la chaîne lourde de toxine, sont ceux d'une toxine botulinique de type B.
- 55 14. Polypeptide selon l'une quelconque des revendications 1 à 9 comprenant une chaîne légère de toxine botulinique ou un fragment ou une variante d'une chaîne légère de toxine botulinique et au moins 100 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique.
15. Polypeptide selon la revendication 14 comprenant une chaîne légère de toxine botulinique de type B, ou un fragment ou un variant de celle-ci, et 107 acides aminés d'extrémité N-terminale d'une chaîne lourde de toxine botulinique de type B.
16. Polypeptide selon la revendication 11 ou la revendication 12 comprenant au moins 423 des acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type A.
17. Polypeptide selon la revendication 16 comprenant une chaîne légère de toxine botulinique de type A et 423 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type A.
18. Polypeptide selon la revendication 16 comprenant un variant de chaîne légère de toxine botulinique de type A dans lequel le résidu 2 est un glutamate, le résidu 26 est une lysine et le résidu 27 est une tyrosine, et 423 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type A.
19. Polypeptide selon la revendication 13 comprenant au moins 417 des acides aminés N-terminaux de la chaîne lourde de toxine botulinique de type B.

20. Polypeptide selon la revendication 19 comprenant au moins une chaîne légère de toxine botulinique de type B et 417 acides aminés N-terminaux d'une chaîne lourde de toxine botulinique de type B.
- 5 21. Polypeptide selon l'une quelconque des revendications 10 à 20 dans lequel il manque une portion désignée H_C d'une chaîne lourde de toxine botulinique est absente.
22. Polypeptide selon l'une quelconque des revendications 1 à 4, dans lequel le deuxième domaine n'est pas capable de se lier à des récepteurs de surface cellulaires.
- 10 23. Polypeptide selon l'une quelconque des revendications précédentes comprenant un variant d'une toxine clostridienne et comprenant en outre un site pour le clivage par une enzyme protéolytique, lequel site de clivage n'est pas présent dans la toxine native.
- 15 24. Polypeptide selon la revendication 23 comprenant un variant d'une chaîne légère de toxine clostridienne et comprenant en outre un site pour le clivage par une enzyme protéolytique, lequel site de clivage n'est pas présent dans la chaîne légère de la toxine native.
- 20 25. Polypeptide selon la revendication 23 ou la revendication 24 comprenant un variant d'une portion H_N de chaîne lourde de toxine clostridienne et comprenant en outre un site pour le clivage par une enzyme protéolytique, lequel site de clivage n'est pas présent dans la portion H_N de la chaîne lourde de la toxine native.
- 25 26. Polypeptide selon la revendication 23, 24 ou 25 pouvant être obtenu par modification d'un ADN codant pour le polypeptide de manière à y introduire un ou plusieurs nucléotides codant pour le site de clivage.
- 30 27. Protéine de fusion formée par la fusion de (a) un polypeptide selon l'une quelconque des revendications 1 à 26 avec (b) un deuxième polypeptide qui est un polypeptide ou un oligopeptide qui se lie à une matrice d'affinité de manière à permettre une purification de la protéine de fusion en utilisant ladite matrice.
- 35 28. Protéine de fusion selon la revendication 27, dans laquelle ledit deuxième polypeptide se lie à une colonne de chromatographie.
- 40 29. Protéine de fusion selon la revendication 28, dans laquelle ladite colonne de chromatographie est une matrice d'affinité de sépharose glutathione.
- 45 30. Protéine de fusion selon la revendication 27, 28 ou 29, dans laquelle un site spécifique de clivage par protéase est incorporé entre les premier et deuxième polypeptides, ledit site de clivage par protéase permettant la séparation protéolytique des premier et deuxième polypeptides.
- 50 31. Composition comprenant une polypeptide selon l'un quelconque des revendications précédentes, ladite composition étant non toxique *in vivo*.
- 55 32. Composition selon la revendication 31 ou polypeptide selon l'une quelconque des revendications 1 à 26 ou protéine de fusion selon la revendication 27, 28, 29 ou 30 destinés à une utilisation comme témoin positif dans un dosage de toxine.
33. Composition selon la revendication 31 ou polypeptide selon l'une quelconque des revendications 1 à 26 ou protéine de fusion selon la revendication 27, 28, 29 ou 30 destinés à une utilisation comme vaccin contre une toxine clostridienne.
34. Composition selon la revendication 31 ou polypeptide selon l'une quelconque des revendications 1 à 26 ou protéine de fusion selon la revendication 27, 28, 29 ou 30 destinés à une utilisation *in vivo*.
35. Composition pharmaceutique comprenant une composition selon la revendication 31, un polypeptide selon l'une quelconque des revendications 1 à 26, ou une protéine de fusion selon la revendication 27, 28, 29 ou 30, en combinaison avec un support physiologiquement acceptable.
36. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon l'une quelconque des revendications 1 à 30.

EP 0 939 818 B1

37. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon la revendication 36 et comprenant des nucléotides codant pour des résidus 1 à 448 d'une chaîne légère de toxine botulinique de type A.
- 5 38. Acide nucléique selon la revendication 36 ou la revendication 37 comprenant des nucléotides codant pour des résidus 1 à 423 d'un domaine H_N d'une chaîne lourde de toxine botulinique de type A.
39. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon la revendication 36 et comprenant des nucléotides codant pour des résidus 1 à 470 d'une chaîne légère de toxine botulinique de type B.
- 10 40. Acide nucléique codant pour un polypeptide ou une protéine de fusion selon la revendication 36 ou la revendication 39 comprenant des nucléotides codant pour des résidus 1 à 417 d'un domaine H_N de chaîne lourde de toxine botulinique de type B.
- 15 41. Acide nucléique selon l'une quelconque des revendications 36 à 40 comprenant des nucléotides codant pour un site de clivage par endonucléase de restriction non présent dans la séquence de toxine clostridienne native.
- 20 42. Acide nucléique selon la revendication 41 susceptible d'être obtenu par modification d'un nucléotide codant pour un polypeptide ou une protéine de fusion selon l'une quelconque des revendications 1 à 30 de sorte à introduire ledit site de clivage.
43. ADN selon l'une quelconque des revendications 36 à 42.
44. ADN choisi parmi SEQ ID Nos 1, 3, 5, 9, 11, 13, 15, 17, 19, 21, 27 et 28.
- 25 45. Procédé de fabrication d'un polypeptide selon l'une quelconque des revendications 1 à 26 comprenant l'expression dans une cellule hôte d'un acide nucléique selon l'une quelconque des revendications 36 à 44 et la récupération polypeptide.
- 30 46. Procédé de fabrication d'un polypeptide selon l'une quelconque des revendications 1 à 26 comprenant l'expression dans une cellule hôte d'un acide nucléique codant pour une protéine de fusion selon la revendication 27, 28, 29 ou 30, la purification de la protéine de fusion par élution de la protéine de fusion à travers une matrice d'affinité adaptée pour retenir la protéine de fusion et par élution à travers ladite matrice d'un ligand adapté pour déplacer la protéine de fusion, et la récupération la protéine de fusion.
- 35 47. Procédé de fabrication selon les revendications 45 ou 46, dans lequel l'acide nucléique est de l'ADN.
48. Cellule exprimant un polypeptide ou une protéine de fusion selon l'une quelconque des revendications 1 à 30.

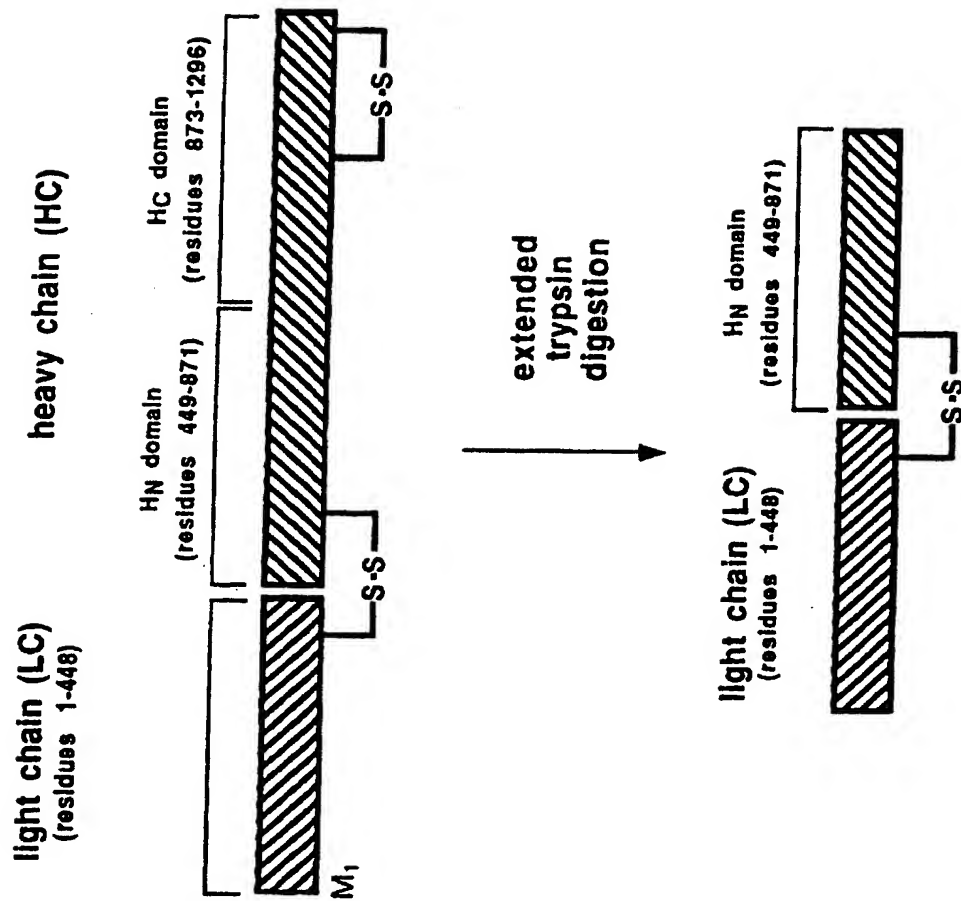


FIG. 1

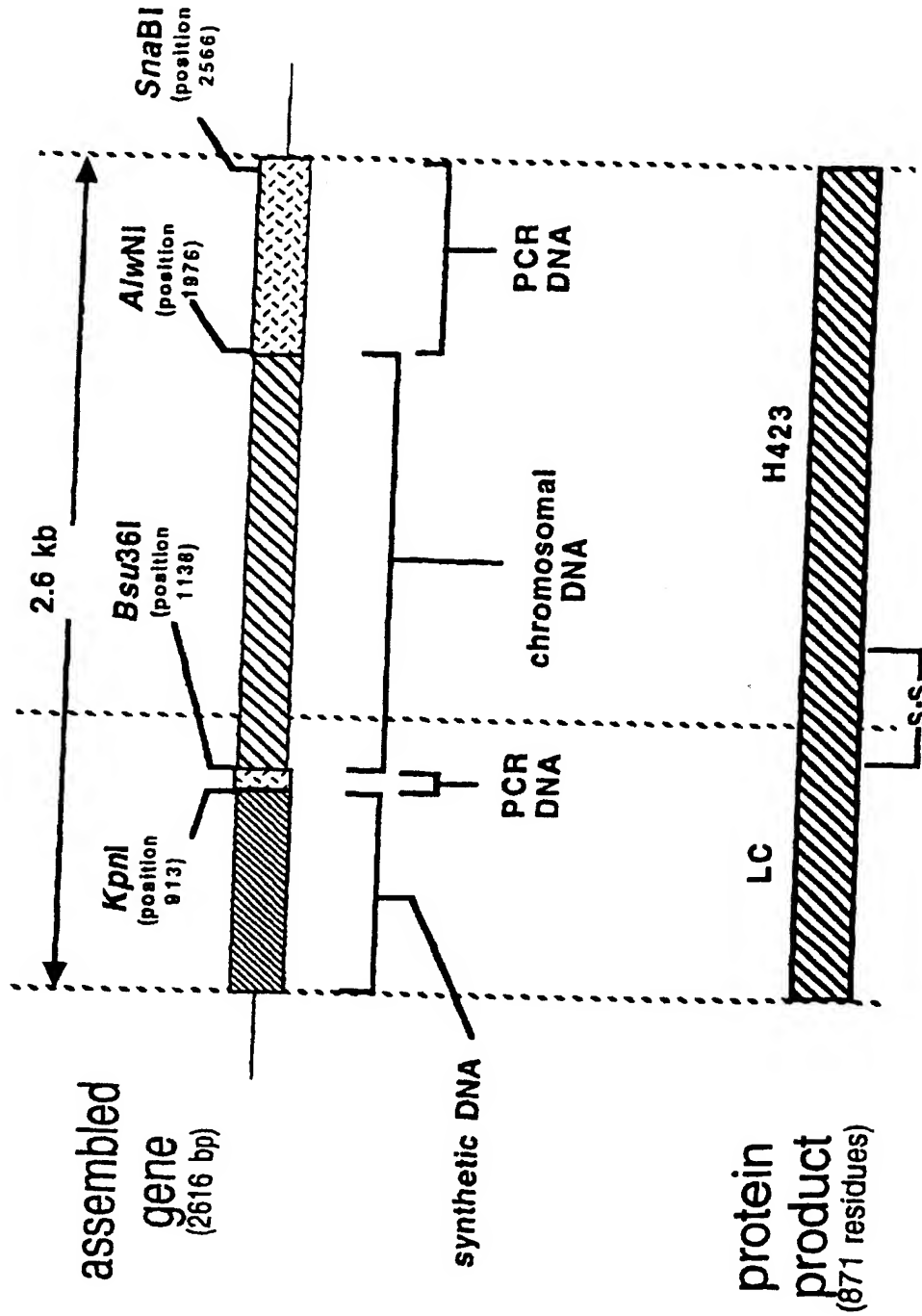


FIG. 2

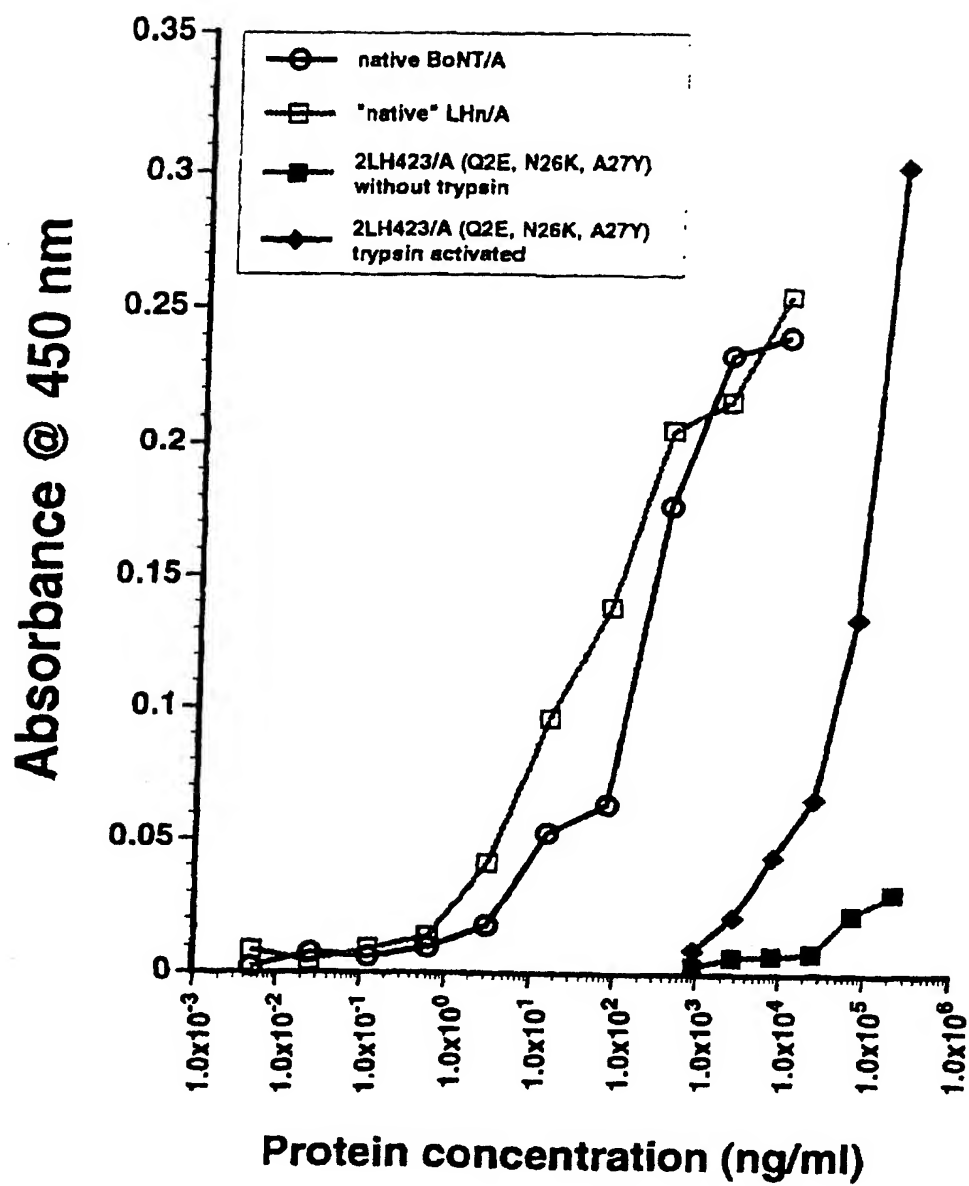


FIG. 3

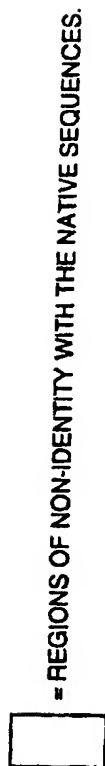


FIG. 4

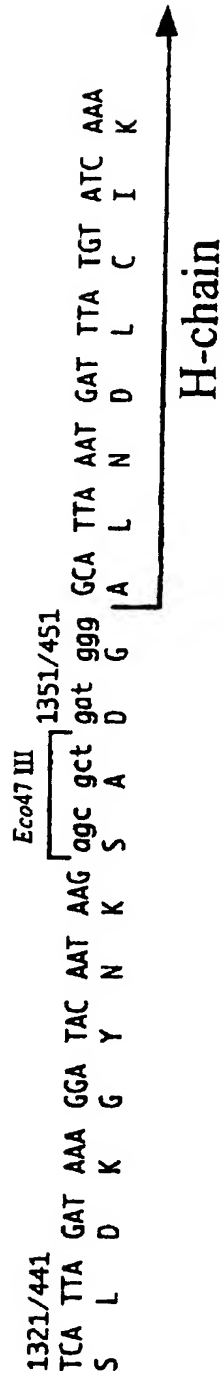


FIG. 5

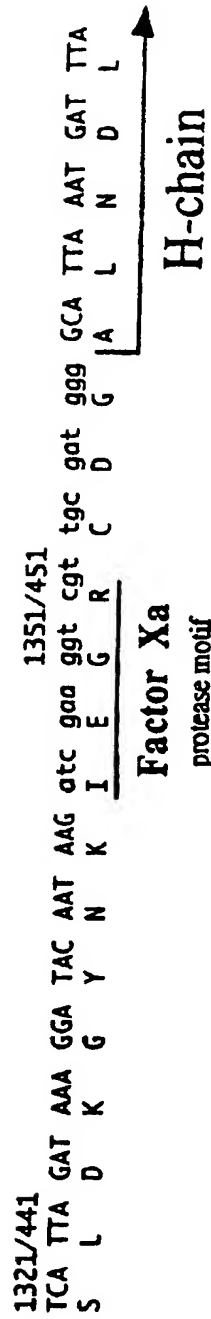


FIG. 6

2587/863 IGF-1
 TAC GTA GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG CCT GGA
 Y V D N Q R L L S T F T E Y I K S R P G
 2647/883
 CCG GAG ACG CTC TGC GGG GCT GAG CTG GTG GAT GCT CTT CAG TTC GTG TGT GGA GAC AGG
 P E T L C G A E L V D A L Q F V C G D R
 2707/903
 GGC TTT TAT TTC AAC AAG CCC ACA GGG TAT GGC TCC AGC AGT CGG AGG GCG CCT CAG ACA
 G F Y F N K P T G Y G S S S R A P Q T
 2767/923
 GGT ATC GTG GAT GAG TGC TGC TTC CCG AGC TGT GAT CTA AGG AGG CTG GAG ATG TAT TGC
 G I V D E C C F R S C D L R L E M Y C
 2827/943
 GCA CCC CTC AAG CCT GCC AAG TCA GCT GAA GCT TAG
 A P L K P A K S A E A stop

FIG. 7

2587/863 CxAl4
 TAC GTA GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG CCT CAA
 Y V D N Q R L L S T F T E Y I K S R P Q
 2647/883
 TCT AAA GTT AAA AGA CAA ATA TTT TCA GGC TAT CAA TCT GAT ATT GAT ACA CAT AAT AGA
 S K V K R Q I F S G Y Q S D I D T H N R
 2707/903
 ATT AAG GAT GAA TTA TGA
 I K D E L stop

FIG. 8

2587/863 TAC GTA GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCA GGC CTG AAT
 Y V D N Q R L L S T F T E Y I K S G L N
 2617/873
 2647/883 TCC CCG GGT GCA GCT CAT TAT GCG CAA CAC GAT GAA GCC GTA GAC AAC AAA TTC AAC AAA
 S P G A A H Y A Q H D E A V D N K F N K
 2677/893
 2707/903 GAA CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC GAA GAA CAA CGA
 E Q Q N A F Y E I L H L P N L N E E Q R
 2737/913
 2767/923 AAC GCC TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC CAA AGC GCT AAC CTT TTA GCA GAA
 N A F I Q S L K D D P S Q S A N L L A E
 2797/933
 2827/943 GCT AAA AAG CTA AAT GAT GCT CAG GCG CCG AAA GTA GAC AAC AAA TTC AAC AAA GAA CAA
 A K K L N D A Q A P K V D N K F N K E Q
 2857/953
 2887/963 CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC GAA CAA CGA AAC GCC
 Q N A F Y E I L H L P N L N E E Q R N A
 2917/973
 2947/983 TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC CAA AGC GCT AAC CTT AAC GCA GAA GCT AAA
 F I Q S L K D D P S Q S A N L L A E A K
 2977/993
 3007/1003 AAG CTA AAT GAT GCT CAG GCG CCG AAA GTA GAC TAG
 K L N D A Q A P K V D *
 3037/1013

FIG. 9

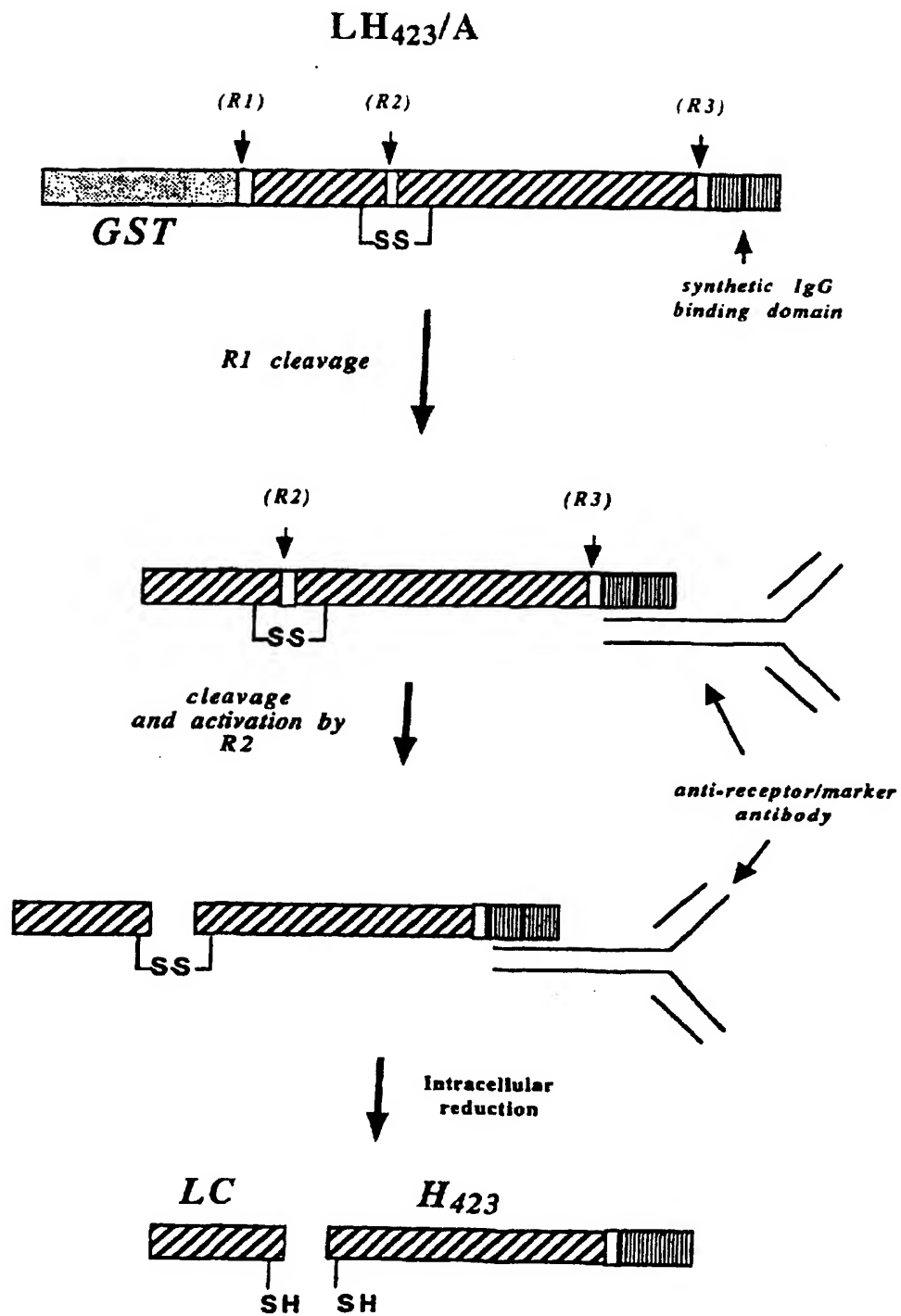


FIG. 10

LH₄₂₃/A

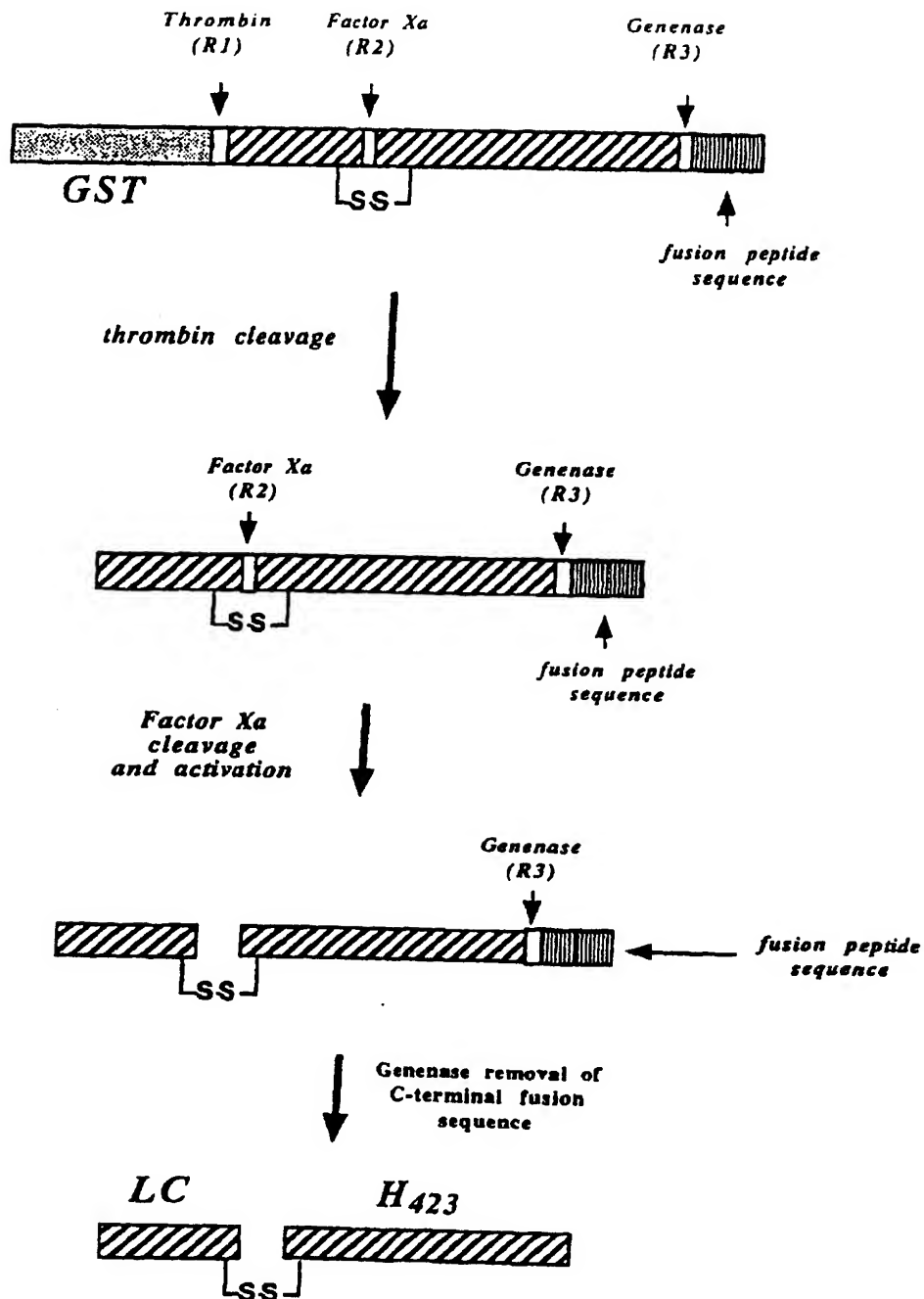


FIG. 11

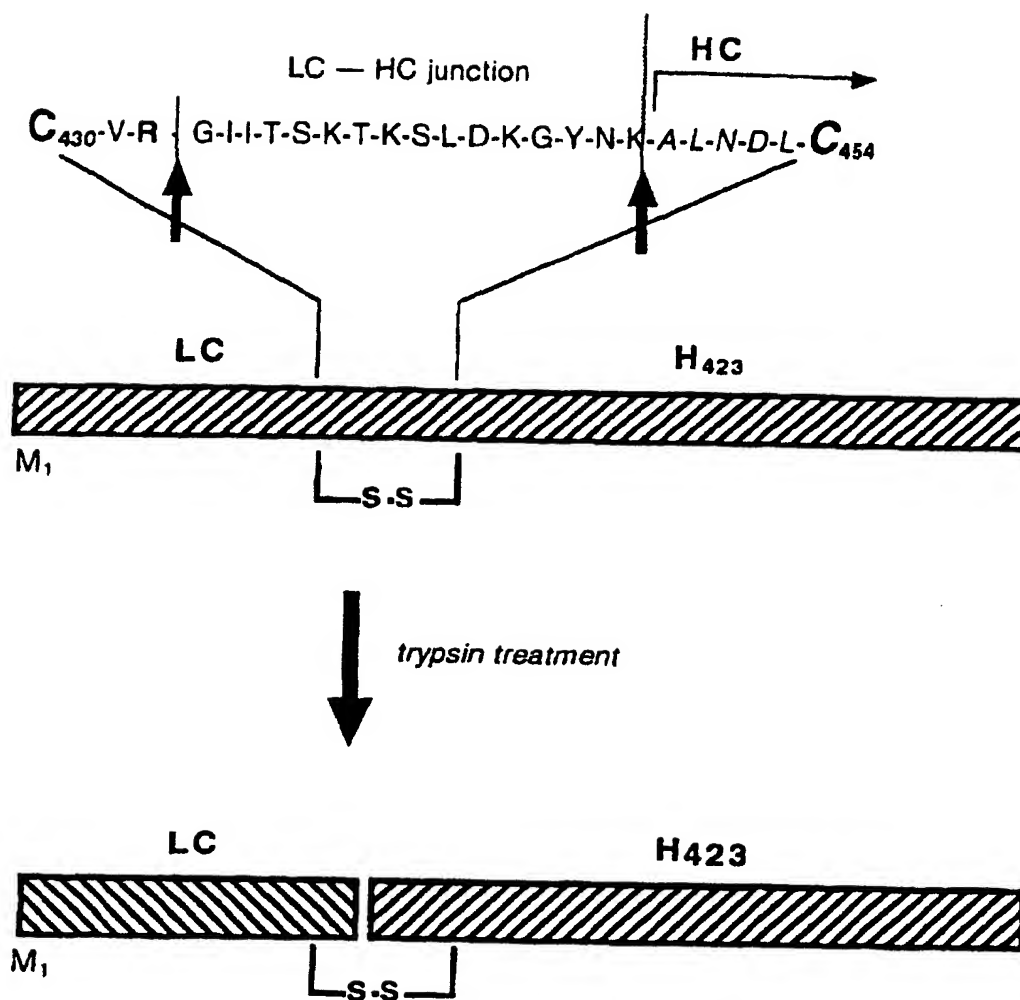


FIG. 12

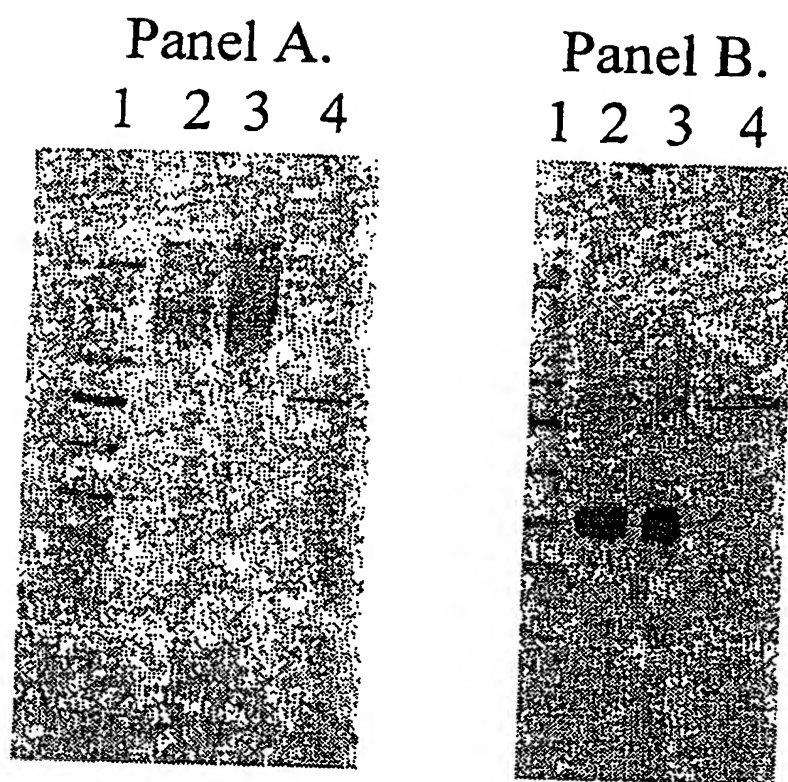


FIG. 13